CORRECTED VERSION*

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 4 : C12Q 1/68 // C07H 21/04 — A1

(11) International Publication Number: WO 89/12697

(43) International Publication Date: 28 December 1989 (28.12.89)

(21) International Application Number: PCT/US89/02731

(22) International Filing Date: 21 June 1989 (21.06.89)

(30) Priority data:
210,116 22 June 1988 (22.06.88) US

(71) Applicant: THE BOARD OF REGENTS OF THE UNIVERSITY OF WASHINGTON [US/US]; 3755 University May N.E., Seattle, WA 98195 (US).

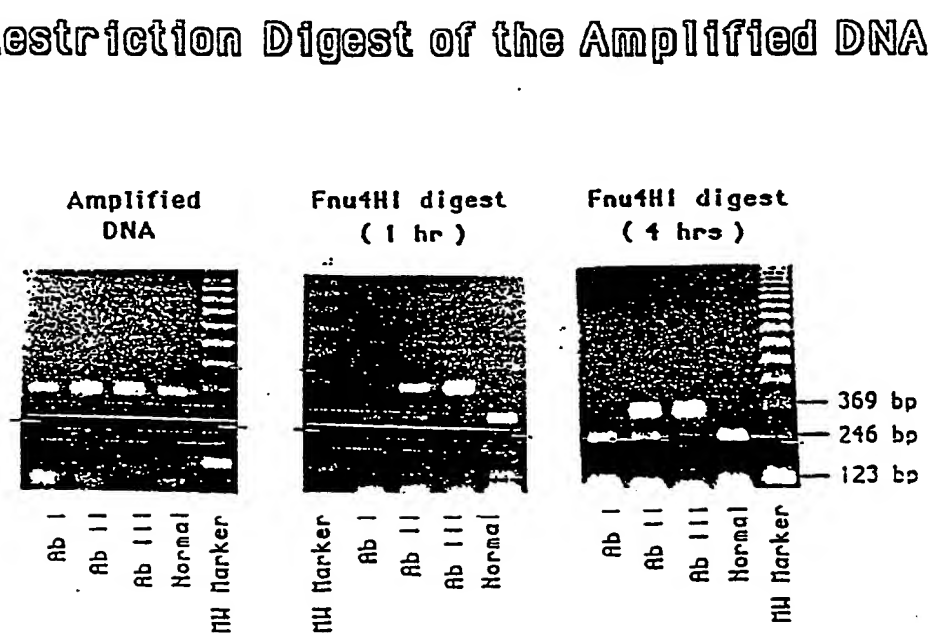
(72) Inventor: ICHINOSE, Akitada ; 6846 40th Avenue N.E., Seattle, WA 98115 (US).

(74) Agents: MAKI, David, J. et al.; Seed and Berry, 6300 Colombia Center, Seattle, WA 98104-7092 (US).

(81) Designated States: AT (European patent), BE (European patent), CH (European patent), DE (European patent), DK, FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHOD FOR DETECTING ABNORMAL GENES



(57) Abstract

Methods for detecting the presence of selected mutations, such as the Thr-601 mutation and the Phe-355 mutation, in the plasminogen of a patient are disclosed. The methods include exposing amplified genomic DNA to a restriction endonuclease capable of differentially cleaving mutant and wild-type plasminogen DNA sequences, and analyzing the exposed DNA to detect the presence or absence of cleavage fragments diagnostic for the selected mutation. Diagnostic kits for the rapid detection of the selected mutation are also disclosed.

* (Referred to in PCT Gazette No. 04/1990, Section II)

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FI	Finland	ML	Mali
AU	Australia	FR	France	MR	Mauritania
BB	Barbados	GA	Gabon	MW	Malawi
BE	Belgium	GB	United Kingdom	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IT	Italy	RO	Romania
BJ	Benin	JP	Japan	SD	Sudan
BR	Brazil	KP	Democratic People's Republic of Korea	SE	Sweden
CF	Central African Republic			SN	Senegal
CG	Congo	KR	Republic of Korea	SU	Soviet Union
CH	Switzerland	LI	Liechtenstein	TD	Chad
CM	Cameroon	LK	Sri Lanka	TG	Togo
DE	Germany, Federal Republic of	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		
ES	Spain	MG	Madagascar		

METHOD FOR DETECTING ABNORMAL GENES

Technical Field

The present invention is related generally to the detection of abnormal genes. More specifically, the invention provides methods for detecting the presence of abnormal plasminogen genes, such as a gene encoding Thr-601 plasminogen or a gene encoding Phe-355 plasminogen.

Background of the Invention

In order to understand the mechanisms and genetics of human diseases, it is important to identify DNA and protein markers that indicate the presence of genetic defects in populations and families. For example, deficiencies in protein C, protein S, antithrombin III, heparin co-factor II, tissue-type plasminogen activator and plasminogen have been identified as the cause or at least part of the cause of a predisposition for thrombosis in some patients with hereditary thrombophilia (for review, see Bauer and Rosenberg, Blood 70:343-350, 1987, and Mannucci and Tripodi, Thromb. Haemostas. 57:247-251, 1987).

Plasminogen is a single-chain proenzyme that is converted to an active two-chain form (consisting of an A and a B chain connected by two disulfide bonds), called plasmin, by activators such as tissue-type plasminogen activator, urokinase, and streptokinase. Plasmin digests fibrin clots to form soluble fibrin degradation products. In addition, plasmin is thought to play an important role in various biological reactions, such as inflammation, tissue development and remodeling, processing other molecules, etc.

The primary structure of plasminogen (790 amino acid residues) was established by Sottrup-Jensen et al. (Prog. Chem. Fibrinol. Thrombol. 3:191-209, 1978). This

amino acid sequence has been confirmed by cDNA sequencing (Malinowski et al., Biochemistry 23:4243-4250, 1984 and Forsgren et al., FEBS Lett. 213:254-260, 1987), which indicated the presence of an additional Ile residue at position 65. Accordingly, plasminogen contains 791 amino acids (See Figure 1). The A chain of the molecule consists of the activation peptide (77 amino acid residues) and five disulfide bond-folded structures called "kringles" (about 90 residues each). The B chain contains the activation site (between Arg-561 and Val-562), the active site His-603 residue region, the active site Asp-646 residue region, the region which is linked to the heavy chain by a disulfide bond, the active site Ser-741 residue region, and the C-terminus (amino acid numbers used herein refer to the sequence shown in Figure 1). The first kringle structure (K1) in the A chain of plasminogen is responsible for its binding to fibrin (Thorsen et al., Biochim. Biophys. Acta. 668:377-387, 1981). The B chain of plasminogen carries all three active sites essential for catalytic function as a serine protease.

There are at least several genes in the human genome that are homologous to that of plasminogen, such as apolipoprotein(a) (McLean et al., Nature 330:132-137, 1987). Apolipoprotein(a) contains 37 copies of plasminogen kringle 4 and one copy of plasminogen kringle 5. It also contains a serine protease domain that is highly homologous with the B chain of plasminogen.

Several cases of a molecular abnormality of plasminogen in association with a complication of thrombosis have been reported (Aoki et al., J. Clin. Invest. 61:1186-1195, 1978; Kazama et al., Thromb. Res. 21:517-522, 1981; Wohl et al., Thromb. Haemostas. 48:146-152, 1982; Soria et al., Thromb. Res. 32:229-238, 1982 and Scharrar et al., Thromb. Hemostas. 55:396-401, 1986). These abnormalities have been found most frequently in Japan, but have also been reported in other populations. By an analysis of the plasminogen molecules from these patients, it has been

shown that an amino acid substitution of Thr for Ala-601 in the B chain results in the generation of an inactive plasmin molecule (Sakata and Aoki, J. Biol. Chem. 255:5442-5447, 1980; Miyata et al., Proc. Natl. Acad. Sci. USA 79:6132-6136, 1982; Miyata et al., J. Biochem. 96:227-287, 1984). However, the nature of the underlying abnormality at the DNA level has not heretofore been determined, and other plasminogen disorders have not been characterized.

Since plasminogen is the key enzyme in the fibrinolytic system, responsible for removing fibrin clots from circulation, individuals with abnormal plasminogen or a plasminogen deficiency develop thrombosis. Given the gene frequency of approximately 0.02 among Japanese, the expected number of homozygotes with the Thr-601 plasminogen variant is calculated to be about 50,000 in Japan (population of approximately 125 million). A few homozygotes have been found; however, the homozygous condition is expected to be lethal in most cases. In heterozygotes, the reduced plasminogen activity in plasma seems to be insufficient to prevent thrombosis, which may develop after trauma and is manifested as deep vein thrombosis, thrombophlebitis or pulmonary embolism.

Conventional biological assays for plasminogen activity and antigen concentration do not accurately identify the molecular basis of thrombosis, because plasminogen can be decreased in several acquired disease states, such as liver dysfunction and disseminated intravascular coagulation, or by thrombolytic therapy using plasminogen activators. Because proper therapy is dictated by the nature of the underlying condition, it is important to make a definitive diagnosis in the case of a genetic molecular abnormality. An additional complication in diagnosing plasminogen-related disorders arises from the high degree of homology between plasminogen and apolipoprotein(a). This homology makes it difficult to distinguish between DNA sequences encoding the two proteins.

Previously described methods of identifying the presence of the Thr-601 plasminogen mutation are not well suited to clinical use. Miyata et al. (Proc. Natl. Acad. Sci. USA 79:6132-6136, 1982) used proteolytic digestion of plasminogen and amino acid sequence analysis of the resultant peptides to characterize the mutation. Aoki et al. (Biochemical Genetics 22:871-881, 1984) utilized electrofocusing, zymography and immunofixation of neuraminidase-treated plasminogen. The entire procedure required four or more days to perform.

There is therefore a need in the art for improved methods of detecting the presence of mutations in the plasminogen gene. Such methods should be technically simple and rapid enough to permit clinical use. The present invention provides such methods for genetic diagnosis at the DNA level and has the additional advantage of not being influenced by the presence of other disease conditions.

Disclosure of the Invention

Briefly stated, the present invention is directed toward methods for detecting the presence of a mutation in the plasminogen gene of a patient. In one aspect of the present invention, the method comprises (a) amplifying a portion of genomic DNA from a patient, the portion including a predetermined exon comprising the site of a selected mutation and at least 14 base pairs of each of two intron sequences flanking the exon; (b) exposing the amplified DNA to a restriction endonuclease capable of differentially cleaving DNA having the selected mutation and wild-type plasminogen DNA, under conditions suitable for activity of the endonuclease; and (c) analyzing the exposed DNA to detect the presence or absence of cleavage fragments diagnostic for the selected mutation. Within preferred embodiments, the selected mutation is the Phe-355 mutation or the Thr-601 mutation. The method may also include,

prior to the step of amplifying, isolating genomic DNA from the patient.

Within a related aspect of the present invention, a method of detecting the presence of a mutation in the plasminogen gene of a patient is disclosed, wherein the method generally comprises (a) denaturing genomic DNA from the patient; (b) annealing the denatured genomic DNA to a pair of oligonucleotide primers, wherein the first primer is complementary to a first sequence of at least about fifteen consecutive nucleotides of a first intron on the coding strand of the genomic DNA, and wherein the second primer is complementary to a second sequence of at least about fifteen consecutive nucleotides of a second intron on the noncoding strand of the genomic DNA, the introns flanking the exon comprising the site of a selected mutation; (c) extending the annealed primers to produce double-stranded DNA fragments, the fragments including the site of the selected mutation; (d) denaturing the double-stranded DNA fragments; (e) annealing the denatured DNA fragments to the pair of oligonucleotide primers and extending the annealed primers to produce selectively amplified DNA; (f) exposing the selectively amplified DNA to a restriction endonuclease capable of differentially cleaving DNA having the selected mutation and wild-type plasminogen DNA, under conditions suitable for activity of the endonuclease; and (g) analyzing the exposed DNA to detect the presence or absence of cleavage fragments diagnostic for the selected mutation, wherein the selected mutation is the Phe-355 mutation or the Thr-601 mutation. Within a preferred embodiment, the primers are extended using Taq DNA polymerase.

Within another aspect of the present invention, a diagnostic kit for the rapid detection of the Thr-601 mutation in the plasminogen gene of a patient is disclosed. The kit includes, within suitable compartments: a pair of oligonucleotide primers, the first primer being complementary to a first sequence of at least about fifteen consecu-

tive nucleotides of an intron on the coding strand of genomic DNA from a patient, the second primer being complementary to a second sequence of at least about fifteen consecutive nucleotides of a second intron on the noncoding strand of the genomic DNA, the introns flanking the exon coding for amino acid residue 601 of plasminogen; Taq DNA polymerase; control DNA; a restriction endonuclease capable of differentially cleaving Ala-601 plasminogen DNA and Thr-601 plasminogen DNA; and suitable buffers.

Within yet another aspect of the present invention, a diagnostic kit for the rapid detection of the Phe-355 mutation in the plasminogen gene of a patient is provided. The kit comprises, contained within suitable compartments, (a) a pair of oligonucleotide primers, the first primer being complementary to a first sequence of at least about fifteen consecutive nucleotides of an intron on the coding strand of genomic DNA from a patient, the second primer being complementary to a second sequence of at least about fifteen consecutive nucleotides of a second intron on the noncoding strand of the genomic DNA, the introns flanking the exon coding for amino acid 355 of plasminogen; (b) Taq DNA polymerase; (c) control DNA; (d) a restriction endonuclease capable of differentially cleaving Val-355 plasminogen DNA and Phe-355 plasminogen DNA; and (e) suitable buffers.

These and other aspects of the present invention will become evident upon reference to the following detailed description and attached drawings.

Brief Description of the Drawings

Figure 1 illustrates the cDNA sequence and amino acid sequence of plasminogen. The positions of certain restriction enzyme recognition sites are shown. Numbers in the left margin refer to nucleotide positions. Numbers above the sequence refer to amino acid positions.

Figure 2 illustrates portions of the sequence of the normal human plasminogen gene. N indicates an undeter-

mined nucleotide. Arrows indicate exon-intron boundaries. Exon sequences are underlined and labeled with Roman numerals. The 5' end of exon I and the 3' end of exon XIX were not determined; the 3' end of exon XIX is shown at the proposed polyadenylation signal. The partial gene sequence is presented in 10 sections, labeled a through j, showing: a, exon I and adjacent intron sequences; b, exons II and III and adjacent intron sequences; c, exon IV and adjacent intron sequences; d, exon V and adjacent intron sequences; e, exon VI and adjacent intron sequences; f, 10,000 base pairs comprising exons VII, VIII, IX and X; g, 10,000 base pairs comprising exons XI, XII and XIII; h, 10,000 base pairs comprising exons XIV, XV, XVI and XVII; i, intron sequence (4473 bp); and j, exons XVIII and XIX with adjacent intron sequences. Nucleotides in each of sections a through j are independently numbered as designated in the right margin, beginning with 1.

Figure 3 illustrates a portion of the genomic DNA sequence encoding plasminogen and the sequences of two sets of oligonucleotide primers (designated A39, 1A, 10A and 11A) used to selectively amplify a portion of the genomic DNA. The locations of certain restriction enzyme recognition sites are indicated.

Figure 4 shows the results of a Fnu 4HI digest of selectively amplified genomic DNAs from three unrelated patients with abnormal plasminogen and a normal individual. The molecular weight marker is a 123-bp ladder obtained from Bethesda Research Laboratories. AbI, II and III refer to samples from abnormal patients I, II and III, respectively.

Best Mode for Carrying Out the Invention

Prior to setting forth the invention, it may be useful to define certain terms used herein.

Selectively amplifying: The process of increasing the copy number of a preselected DNA sequence or

fragment relative to the copy number of other sequences or fragments in a sample.

Differentially cleaving: Cleaving a first sequence or set of sequences but not cleaving a second sequence or set of sequences. Restriction endonucleases differentially cleave DNA sequences due to their ability to specifically recognize short stretches of paired bases, frequently palindromic sequences of four to six base pairs. Cleavage may occur within the recognition sequence or at some specific distance away from the recognition sequence.

Site of the selected mutation: The position in a gene at which a mutation is known to occur, regardless of whether that particular allele carries the mutant or wild-type sequence at the site.

As noted above, reduced plasminogen activity can lead to thrombotic episodes. Also as noted above, such a reduction in activity can result from a variety of causes, including genetic abnormalities. Practical methods of clinical screening for genetic abnormalities in plasminogen have heretofore been unavailable.

The present invention provides methods useful in diagnosing cases of thrombosis, in genetic screening and in prenatal diagnosis. The methods are simple, rapid, and do not require the use of radioactive isotopes, so are particularly useful in many clinical laboratories that lack in the special facilities necessary for handling radioisotopes.

The present invention is related, in part, to the elucidation of the human plasminogen gene sequence, portions of which are shown in Figures 2 and 3. Knowledge of this sequence has permitted the design of oligonucleotide primers that may be used to selectively amplify those portions of the gene encoding amino acid residue 601 or amino acid residue 355. In a similar manner, other abnormal plasminogen gene sequences may be analyzed, allowing those skilled in the art to selectively amplify exons comprising sites of other selected mutations.

The methods of the present invention are applied to genomic DNA samples from a patient. In one embodiment, the genomic DNA is first isolated, using conventional procedures. A convenient source of isolated genomic DNA is leukocytes, which may be readily obtained from a small (e.g., 10 ml) blood sample. Other cell types may also be used. DNA may be isolated from leukocytes using the technique of Bell et al. (Proc. Natl. Acad. Sci. USA 78:5759-5763, 1981). Briefly, blood is collected in the presence of an anticoagulant, the cells are lysed, and the nuclei are collected. The nuclei are then treated with sodium dodecyl sulfate and proteinase K and the DNA is extracted from the mixture with phenol/chloroform/isoamyl alcohol. The DNA is then precipitated and resuspended in a suitable buffer, such as 10 mM Tris-HCl (pH 7.5), 1 mM EDTA. Alternatively, by using the method disclosed disclosed by Kogan et al. (New Eng. J. Med. 317:985-990, 1987), the methods of the present invention may be applied directly to tissue samples, without the need to isolate the DNA. For example, chorionic villus samples can be screened directly by disrupting the tissue by vortexing in a solution of 0.1M NaOH, 2M NaCl, 0.5% SDS. The sample is then boiled for two minutes, centrifuged, and an aliquot is taken for amplification. This facilitates the application of these methods to prenatal diagnosis of the plasminogen abnormality.

Genomic DNA (either isolated or in the form of a suitable tissue sample) is then selectively amplified to provide a high copy number of the desired portion of the plasminogen gene (e.g., the portion encoding amino acid residue 601 or the portion encoding amino acid residue 355). Preferably, a sequence of approximately 200-1,000, most preferably about 300-400, base pairs is selectively amplified. In a preferred embodiment, the exon encoding amino acid 601 and portions of the intron sequences flanking this exon are selectively amplified. Similarly, the exon encoding amino acid 355 and portions of the flanking introns may be selectively amplified. A preferred method

of amplification is the polymerase chain reaction, described by Mullis (U.S. Patent Nos. 4,683,202 and 4,683,195). Briefly, the genomic DNA is denatured to separate the coding and noncoding strands. Denaturation is preferably accomplished by heat treatment of the DNA, generally treatment at about 80°C-105°C for about one to ten minutes, although enzymatic denaturation may also be used. Most preferably, the DNA is heated at about 93°C for one minute. The denatured DNA is then combined with a molar excess of a pair of oligonucleotide primers under conditions which allow the DNA strands to anneal to the primers (e.g., 60°C for one to three minutes, preferably about two minutes). Preferably, each primer is used at a concentration of about 1μM for amplification of one microgram of genomic DNA. Suitable results may be obtained with 5μg of primer per μg of target DNA. One of the primers is complementary to a sequence on the coding strand and the second primer is complementary to a sequence on the noncoding strand, the sequences flanking the region to be amplified. "Sequences flanking the region to be amplified" include exon sequences, sequences of introns immediately adjacent to the exon to be amplified and sequences of other introns, so long as the amplified region includes the site of the selected mutation. The flanking sequences should be selected so as to provide an amplified portion of the gene within the size limits noted above. Although 100% complementarity is not required, a high degree of complementarity of primer and genomic DNA is advantageous in that it results in high specificity and efficiency of amplification. For use within the present invention, the primers must be sufficiently complementary to hybridize with their respective strands on the genomic DNA. The annealed primers are enzymatically extended using a DNA polymerase and all four deoxyribonucleotide triphosphates (dNTP's). Suitable polymerases include E. coli DNA polymerase I, the Klenow fragment of E. coli DNA polymerase I, Taq DNA polymerase, and T4 DNA polymerase. Taq DNA polymerase (Saiki et al.,

Science 239:487-491, 1988) is particularly preferred. The reaction mixture is incubated under conditions of time and temperature suitable for the activity of the polymerase. When using the Taq DNA polymerase the mixture is incubated at about 70°C ± 10°C for approximately three minutes. As will be appreciated by one skilled in the art, the exact time and temperature will be determined by the melting point of the annealed DNA. The resulting extension products are separated from the original DNA strands, preferably by heat denaturation. The annealing, extension and separation steps are then repeated, preferably about 25 to 30 times, until the desired degree of amplification is obtained. At that time, the final separation step is omitted, and double-stranded DNA is isolated. In general, it is preferred to add the primers and dNTP's at the beginning of the amplification reaction in sufficient quantity to allow full amplification to occur without the need to add additional reagents during the course of the reaction series. The use of Taq DNA polymerase facilitates such a process, as this heat-stable enzyme is not inactivated by the heat denaturation steps and the reaction need not be interrupted for the addition of more polymerase.

As noted above, oligonucleotide primers for use in the polymerase chain reaction are constructed to be complementary to sequences flanking an exon comprising the site of a selected mutation, such as the exon containing the codon for amino acid 601 or the exon containing the codon for amino acid 355. A first primer is designed to be complementary to a sequence on the coding strand, and a second primer is complementary to a sequence on the noncoding strand of the DNA. Preferably, the primers will be complementary to intron sequences because intron sequences will exhibit the least amount of intergene homology. The primers are preferably at least about 15-20 bases in length, more preferably at least about 25 bases in length. Primers shorter than about 20 bases will often have reduced specificity, and may anneal to and amplify

unwanted sequences. Primers are preferably less than 50 bases in length, more preferably less than about 30 bases in length. Longer primers may self-anneal or their use may lead to reduced specificity.

Within the present invention, alternative methods of DNA amplification may also be used. For example, a genomic library may be prepared by digesting genomic DNA from a patient and cloning the resultant DNA fragments into a suitable vector (e.g., plasmid, cosmid or bacteriophage). The library is then amplified by conventional methods, and plasminogen-encoding clones are screened for the presence of the mutation.

The amplified DNA is then incubated with a restriction endonuclease which is capable of differentially cleaving normal and abnormal plasminogen DNA. Suitable restriction endonucleases for identification of the Thr-601 mutation include Fnu 4HI and Bbv I. Endonucleases suitable for identification of the Phe-355 mutation include Ava II, Bam NxI, Cau I (Bingham and Darbyshire, Gene 18:87-91, 1982; Molemans et al., Gene 18:93-96, 1982), Hgi BI, Hgi CII, Hgi EI and Sau 96I. However, the invention is not limited to the use of particular enzymes, but is intended to include the use of other suitable enzymes which may from time to time become available. Restriction endonucleases are commercially available from, for example, New England Biolabs (Beverly, Mass.), Bethesda Research Laboratories (Gaithersburg, Md.) and other suppliers. The amplified DNA is incubated with the endonuclease under conditions of time, temperature and buffer composition suitable for the activity of the endonuclease. Such conditions are generally specified by the supplier.

Following exposure to the restriction endonuclease, the DNA sample is analyzed to detect the presence or absence of cleavage fragments diagnostic for the selected mutation, for example by electrophoretic separation of DNA fragments. In a preferred embodiment, the DNA is electrophoresed on an agarose gel containing ethidium

bromide. Endonuclease Fnu 4HI cleaves the normal plasminogen sequence at the codon for Ala-601. The presence of the Thr-601 mutation prevents this cleavage, resulting in no change in fragment size following exposure to the enzyme. Priming in the introns flanking the codon for amino acid 601 as disclosed in more detail below resulted in amplification of a ~340 bp fragment. The normal sequence could be cleaved by Fnu 4HI to yield fragments of about 240 bp and 100 bp. Also, as discussed in more detail below, the mutation of Val-355 to Phe can be detected by amplifying a ~390 bp fragment, digesting the amplified DNA with Ava II and analyzing the digested DNA. The Phe-355 mutation results in the presence of a 360 bp fragment, which is not present in the Ava II digest of wild-type DNA.

The methods described herein are well suited to clinical use. In particular, the combination of the polymerase chain reaction and restriction analysis can be used to diagnose the specific plasminogen abnormality at the DNA level in a rapid and straightforward manner. Partial purification of genomic DNA from leukocytes takes several hours, and amplification by the polymerase chain reaction takes about three hours. Restriction digestion of the amplified DNA and its analysis on agarose gels require about one hour or less each. Therefore, the entire diagnostic procedure can be performed in a single day.

As briefly described above, suitable kits for diagnosing these plasminogen mutations contain oligonucleotide primers, Taq DNA polymerase, an appropriate restriction enzyme, buffers, and normal (control) DNA in appropriate packaging.

The following examples are offered by way of illustration, and not by way of limitation.

EXPERIMENTAL

Taq DNA polymerase was obtained from New England Biolabs and The Perkin Elmer Corporation (Norwalk, Conn.).

Restriction endonucleases and T4 DNA ligase were purchased from Bethesda Research Laboratories (Gaithersburg, Md.) or New England Biolabs. The Klenow fragment of Escherichia coli DNA polymerase, bacterial alkaline phosphatase, ATP, deoxynucleotides, dideoxynucleotides, M13mp18, and M13mp19 were supplied by Bethesda Research Laboratories. dATP[α-^{35}S] was provided by Amersham (Chicago, Ill.).

Oligonucleotides were synthesized using a nucleotide synthesizer (Applied Biosystems, Foster City, Calif.) and kindly supplied by Drs. Patrick S.H. Chou, Yim Foon Lee and Jeff Harris.

Example 1

Leukocyte genomic DNA samples were obtained from three unrelated Japanese patients with abnormal plasminogen (named abnormal I, II and III, respectively), a daughter of abnormal III (abnormal III-2) and three unrelated normal American white individuals. Abnormals I, II and III-2 had a history of thrombosis, but abnormal III did not. The plasma of abnormal I had a trace of plasminogen activity in spite of a normal plasminogen antigen concentration, and the plasma from the mother and a sister of abnormal I showed a 50% reduction in enzymatic activity of plasminogen. Accordingly, abnormal I is a homozygote of a nonfunctional plasminogen variant. Abnormal II is a heterozygote of a plasminogen variant, since the plasminogen in the plasma of the patient and his two daughters has about half of the specific activity (activity per antigen) of normal plasminogen. Abnormal III is a homozygote of the plasminogen variant named PLG B (Nishimukai et al., Hum. Hered. 36:137-142, 1986) as determined by isoelectric focusing. Abnormal III-2 is a heterozygote of PLG B with a normal plasminogen concentration and half of normal specific activity.

Genomic DNA samples were prepared from the leukocytes of the patients with abnormal plasminogen and from normal individuals by the method of Bell et al. (ibid.). Typically, 10-40 ml of blood is collected in citrate buffer.

Ten ml of blood is added to 90 ml of 0.32 M sucrose, 10 mM Tris pH7.5, 5mM $MgCl_2$, 1% Triton X-100, and the mixture is incubated at 4°C to lyse the cells. Nuclei are collected by centrifugation at 1,000 x g for 10 minutes and resuspended in 4.5 ml of 0.075 M NaCl, 0.024 M EDTA, pH 8.0. The nuclei are treated with SDS and proteinase K, and the DNA is extracted with chloroform/phenol/isoamyl alcohol, precipitated with ethanol and resuspended in the appropriate buffer.

Nucleotide primers A39 and 1A (Figure 3) for the putative introns N and O flanking the exon coding for the amino acid residue-601 of plasminogen (exon XV) were synthesized for the polymerase chain reaction. These regions were selected because they lie outside the putative exon 15, and upon selective amplification they produce a fragment of a length suitable for analysis by restriction digestion and DNA sequencing. Both the 5'- and 3'-ends were modified to generate convenient restriction sites (Hind III) for cloning directly into the M13 sequencing vector. One µg of genomic DNA was amplified in a 100 µl reaction mixture containing 50 mM KCl, 10 mM Tris (pH 8.4), 2.5 mM $MgCl_2$, each primer (A39 and 1A, Figure 3) at 1 µM, each dNTP at 200 µM, gelatin at 200 µg/ml, and 2.5 units of Taq DNA polymerase (Saiki et al., Science 239:487-491, 1988). The sample was placed in a small Eppendorf tube and overlaid with 100 µl of mineral oil to prevent evaporation. The sample was heated at 93°C for one minute to denature the DNA, cooled to 60°C for two minutes to anneal the primers, and incubated at 70°C for three minutes to extend the annealed primers. The procedure was repeated for a total of 25-30 cycles of amplification. At the end of the last cycle, the sample was incubated at 70°C for 7 minutes to ensure the completion of the final extension step. After precipitation with ethanol and resuspension in 100 µl TE buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA), 5 µl was applied to a 1.5% agarose gel for submerged electrophoresis, and stained with ethidium bromide. A discrete band of

about 340 bp was obtained for each sample, as predicted from the sequence of the gene for normal plasminogen.

The samples from abnormals I, II, III, III-2 and normal individuals were digested with three units of Fnu 4HI endonuclease for one hour or with six units of enzyme for four hours at 37°C. Five microliters of each sample was then applied to a 1.5% agarose gel containing ethidium bromide. The 340 bp fragment of normal DNA was cleaved into two fragments (about 240 and 100 bp), while that of the DNA from abnormal III remained unchanged (Figure 4). The Fnu 4HI digests of the 340 bp fragments from abnormals II and III-2 each showed a mixed pattern of normal DNA and the DNA from abnormal III. In contrast, the DNA from abnormal I was cleaved completely. Prolonged digestion of the samples for four hours with six units of enzyme gave exactly the same results (Figure 4). The amplification and digestion of the genomic DNAs from abnormals I, II, III and III-2 was performed eight, two, three and two times, respectively, and the results obtained were the same in each experiment for each sample. Fnu 4HI recognizes only th GCNGC sequence, suggesting that one or more of these fou nucleotides in the DNAs of abnormals III, III-2, and II i replaced by other nucleotides. Alternatively, a short stretch of nucleotides could be deleted or inserted in the abnormal DNA.

To characterize the mutation(s) at the DNA level, the amplified fragments were sequenced. Since both the 5'- and 3'-end primers were designed to produce double-stranded fragments flanked by Hind III recognition sequences, the amplified 340 bp fragments from normal and abnormal individuals were digested with Hind III and ligated into M13 sequencing vectors cut with Hind III. In order to obtain the DNA sequence coding for the specific region around amino acid residue 601, the amplified DNAs were also digested with Hinc II and Pst I endonucleases. The digested samples were electrophoresed on a 1.5% agarose gel, electroeluted, and dialyzed against 0.1X TBE (1X TBE

is 89 mM Tris-borate, 89 mM boric acid, 20 mM EDTA) overnight. The dialyzed samples were extracted with phenol and chloroform, precipitated with ethanol, resuspended in TE, and finally subcloned into M13mp18 or mp19 in order to obtain discrete overlapping sequences. The DNA sequences of the inserts were then obtained using the dideoxynucleotide method (Sanger et al. Proc. Natl. Acad. Sci. USA 74:5463-5467, 1977) with dATP [α-^{35}S] and buffer gradient gels (Biggin et al. Proc. Natl. Acad. Sci. USA 80:3963-3965, 1983).

The DNA sequences obtained from the three normal individuals included 343 bp. These sequences were the same as expected for the normal gene except for the presence of Hind III sites at both the 5'- and 3'ends. The sequence of the Hinc II-Pst I fragments from the normal DNAs included 205 bp, and was also the same as the established sequence of the normal gene for plasminogen. The actual sequence of the region coding for amino acid 601 (Ala) included ACTGCTGC in the normal gene.

On the other hand, the DNA sequence analysis of both Hind III and Hinc II-Pst I fragments of abnormal III revealed that the gene of abnormal III contained the sequence ACTACTGC. This corresponds to a single base change resulting in the substitution of Thr (ACT) for Ala (GCT). Twenty-three templates from the amplified samples of abnormal III were sequenced and all of them showed the same abnormal sequence (G to A change). No other alterations of nucleotides were found by DNA sequence analysis.

When twelve templates for abnormal II were sequenced, one-half of them showed the same sequence as the normal gene except for a point mutation (T to C) 5 nucleotides prior to the Fnu 4HI site, and the other half had the same abnormal sequence as abnormal III. These results confirmed that abnormal III is a homozygote of a plasminogen variant and that abnormal II is a heterozygote of the same variant allele.

The exon XV DNA sequence of abnormal I was the same as that of the normal gene, indicating that the abnormality in this molecule is in another region.

A second set of primers (designated 10A and 11A in Figure 3), flanked by Eco RI recognition sequences and four additional nucleotides, was used to confirm the results. A band of 360 bp was obtained for each sample as predicted.

Example 2

Plasminogen gene exon X DNA of abnormal I was amplified essentially as described above using primers K4a-5' (5' GTC AGA ATT CTC AGA GGC TAC CGT ACT 3'; coding strand primer) and K4a-3' (5' CTA CGA ATT CTG GCT CTA ACA CAA ATT TCC 3'; noncoding strand primer). The amplified DNA was digested with Eco RI, and the resulting ~390 bp fragment was cloned into an M13 phage vector and sequenced. Sequence analysis revealed the presence of the sequence GTGTTCCAG in six of the templates, as compared to the wild-type sequence GTGGTCCAG. This T for G substitution results in the substitution of a phenylalanine residue for the normal valine residue at amino acid position 355, located several residues upstream of Kringle 4 in the A chain (Figure 1).

DNA samples from normal and abnormal individuals were digested with five units of Ava II endonuclease for one hour at 37°C. The 390 bp DNA fragment from the normal individuals was cleaved into three fragments of approximately 230 bp, 130 bp and 30 bp. DNA samples from abnormal I and two daughters (abnormals I-2 and I-3) and a nephew (abnormal I-4) of abnormal I showed a mixture of 360 bp, 230 bp, 130 bp and 30 bp fragments. These results indicated that the abnormal patients were heterozygous for the Phe-355 mutation. Thus, this mutation can be diagnosed by the presence of a 360 bp Ava II fragment when DNA is selectively amplified using primers K4a-5' and K4a-3'.

In a second series of experiments, DNA from abnormals I, I-2, I-3 and I-4 was amplified using primers K4a-5' and K4a-32 (5' AAA TGA ATT CCT AGG AAG TTG GCT TGA AGC 3'; noncoding strand primer). Digestion of the resulting ~370 bp fragment with Ava II confirmed the loss of an Ava II site in the abnormal DNA, and also confirmed the diagnosis of abnormals I, I-2, I-3 and I-4 as heterozygotes of the Phe-355 mutation.

From the foregoing, it will be appreciated that, although specific embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without deviating from the spirit and scope of the invention. Accordingly, the invention is not limited except as by the appended claims.

Claims

1. A method of detecting the presence of a mutation in the plasminogen gene of a patient, comprising:

amplifying a portion of genomic DNA from the patient, said portion including a predetermined exon comprising the site of a selected mutation and at least 14 base pairs of each of two intron sequences flanking said predetermined exon;

exposing said amplified DNA to a restriction endonuclease capable of differentially cleaving DNA having the selected mutation and wild-type plasminogen DNA under conditions suitable for activity of the endonuclease; and

analyzing the exposed DNA to detect the presence or absence of cleavage fragments diagnostic for the selected mutation.

2. The method of claim 1 wherein the selected mutation is the Phe-355 mutation or the Thr-601 mutation.

3. A method of detecting the presence of a mutation in the plasminogen gene of a patient, comprising:

a. denaturing genomic DNA from the patient;

b. annealing the denatured genomic DNA to a pair of oligonucleotide primers, wherein the first primer is complementary to a first sequence of at least about fifteen consecutive nucleotides of a first intron on the coding strand of the genomic DNA, and wherein the second primer is complementary to a second sequence of at least about fifteen consecutive nucleotides of a second intron on the noncoding strand of the genomic DNA, wherein said introns flank the exon comprising the site of a selected mutation;

c. extending the annealed primers to produce double-stranded DNA fragments, said fragments including the site of the selected mutation;

d. denaturing the double-stranded DNA fragments;
e. annealing the denatured DNA fragments to the pair of oligonucleotide primers and extending the annealed primers to produce selectively amplified DNA;
f. exposing said selectively amplified DNA to a restriction endonuclease capable of differentially cleaving DNA having the selected mutation and wild-type plasminogen DNA, under conditions suitable for activity of the endonuclease; and
g. analyzing the exposed DNA to detect the presence or absence of cleavage fragments diagnostic for the selected mutation, wherein the selected mutation is the Phe-355 mutation or the Thr-601 mutation.

4. The method of claim 3 wherein the primers are extended using Taq DNA polymerase.

5. The method of claim 3 wherein each of said first and second primers is from about twenty to about thirty nucleotides in length, inclusive.

6. The method of claim 3 wherein said selected mutation is the Thr-601 mutation and said first primer includes the sequence CAA TTT AAC TAA AAT TTG AAC TAA AT or TGT ACA ATG GAG CAG AAC AAA.

7. The method of claim 3 wherein said selected mutation is the Thr-601 mutation and said second primer includes the sequence TCA TGT CTA CTA AAA CAC CCG GAC TTA or TCT CCT TTC TGT GTC ATG TCT A.

8. The method of claim 3 wherein said selected mutation is the Phe-355 mutation and said first primer includes the sequence GTC AGA ATT CTC AGA GGC TAC CGT ACT.

9. The method of claim 3 wherein said selected mutation is the Phe-355 mutation and said second primer includes the sequence CTA CGA ATT CTG GCT CTA ACA CAA ATT TCC or AAA TGA ATT CCT AGG AAG TTG GCT TGA AGC.

10. The method of claim 3, further comprising the step of isolating genomic DNA from the patient prior to the step of denaturing the genomic DNA.

11. The method of claim 3 wherein the endonuclease differentiates between G and A in the first position of the codon for amino acid 601 of plasminogen.

12. The method of claim 11 wherein the restriction endonuclease is selected from the group consisting of Fnu 4HI and Bbv I.

13. The method of claim 3 wherein said endonuclease differentiates between G and T in the first position of the codon for amino acid 355 of plasminogen.

14. The method of claim 13 wherein the restriction endonuclease is selected from the group consisting of Ava II and Sau 96I.

15. The method of claim 3 wherein the steps of denaturing comprise heat treatment of the DNA.

16. The method of claim 3 wherein approximately 300-400 bp of genomic DNA is amplified.

17. The method of claim 3 wherein steps d and e are repeated in sequence from about twenty-three to about twenty-eight times prior to step f.

18. A diagnostic kit for the rapid detection of the Thr-601 mutation in the plasminogen gene of a patient, comprising in suitable compartments within the kit:

a pair of oligonucleotide primers, the first primer being complementary to a first sequence of at least about fifteen consecutive nucleotides of an intron on the coding strand of genomic DNA from a patient, the second primer being complementary to a second sequence of at least about fifteen consecutive nucleotides of a second intron on the noncoding strand of the genomic DNA, the introns flanking the exon coding for amino acid residue 601 of plasminogen;

Taq DNA polymerase;

control DNA;

a restriction endonuclease capable of differentially cleaving Ala-601 plasminogen DNA and Thr-601 plasminogen DNA; and

suitable buffers.

19. A diagnostic kit for the rapid detection of the Phe-355 mutation in the plasminogen gene of a patient, comprising in suitable compartments within the kit:

a pair of oligonucleotide primers, the first primer being complementary to a first sequence of at least about fifteen consecutive nucleotides of an intron on the coding strand of genomic DNA from a patient, the second primer being complementary to a second sequence of at least about fifteen consecutive nucleotides of a second intron on the noncoding strand of the genomic DNA, the introns flanking the exon coding for amino acid 355 of plasminogen;

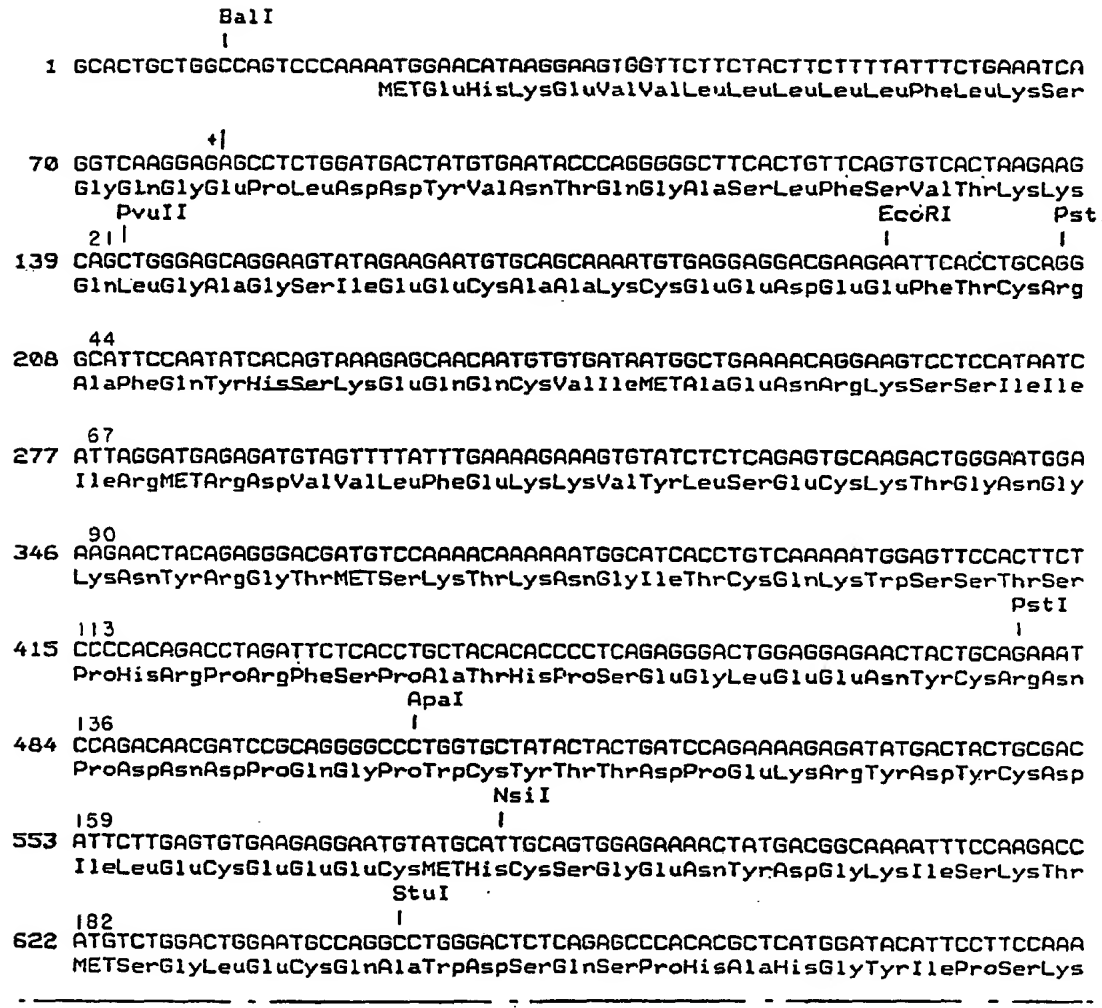
Taq DNA polymerase;

control DNA;

a restriction endonuclease capable of differentially cleaving Val-355 plasminogen DNA and Phe-355 plasminogen DNA; and

suitable buffers.

FIG. 1



```
    205
691 TTTCCAAACAAGAACCTGAAGAAGAATTACTGTCGTAACCCCGAGAGGGAGCTGCGGCCTTGGTGTTTC
    PheProAsnLysAsnLeuLysLysAsnTyrCysArgAsnProGluArgGluLeuArgProTrpCysPhe
                        Eco47III
    228                 |
760 ACCACCGACCCCAACAAGCGCTGGGAACTTTGTGACATCCCCCGCTGCACAACACCTCCACCATCTTCT
    ThrThrAspProAsnLysArgTrpGluLeuCysAspIleProArgCysThrThrProProProSerSer

    251
829 GGTCCCACCTACCAGTGTCTGAAGGGAACAGGTGAAAACTATCGCGGGAATGTGGCTGTTACCGTGTCC
    GlyProThrTyrGlnCysLeuLysGlyThrGlyGluAsnTyrArgGlyAsnValAlaValThrValSer
                          ApaLI
    274                   |
898 GGGCACACCTGTCAGCACTGGAGTGCACAGACCCCTCACACACATAACAGGACACCAGAAAACTTCCCC
    GlyHisThrCysGlnHisTrpSerAlaGlnThrProHisThrHisAsnArgThrProGluAsnPhePro
                                                    ApaINcoI
    297                                             ||
967 TGCAAAAATTTGGATGAAAACTACTGCCGCAATCCTGACGGAAAAAGGGCCCCATGGTGCCATACAACC
    CysLysAsnLeuAspGluAsnTyrCysArgAsnProAspGlyLysArgAlaProTrpCysHisThrThr
                          ScaI
    320                   |
1036 AACAGCCAAGTGCGGTGGGAGTACTGTAAGATACCGTCCTGTGACTCCTCCCCAGTATCCACGGAACAA
    AsnSerGlnValArgTrpGluTyrCysLysIleProSerCysAspSerSerProValSerThrGluGln
                                                    NcoI
    343                                             |
1105 TTGGCTCCCACAGCACCACCTGAGCTAACCCCTGTGGTCCAGGACTGCTACCATGGTGATGGACAGAGC
    LeuAlaProThrAlaProProGluLeuThrProValValGlnAspCysTyrHisGlyAspGlyGlnSer

    366
1174 TACCGAGGCACATCCTCCACCACCACCACAGGAAAGAAGTGTCAGTCTTGGTCATCTATGACACCACAC
    TyrArgGlyThrSerSerThrThrThrThrGlyLysLysCysGlnSerTrpSerSerMETThrProHis
                                                          PstI
    389                                                   |
1243 CGGCACCAGAAGACCCCAGAAAACTACCCAAATGCTGGCCTGACAATGAACTACTGCAGGAATCCAGAT
    ArgHisGlnLysThrProGluAsnTyrProAsnAlaGlyLeuThrMETAsnTyrCysArgAsnProAsp
                                                      ScaI
    412                                               |
1312 GCCGATAAAGGCCCCTGGTGTTTTACCACAGACCCCAGCGTCAGGTGGGAGTACTGCAACCTGAAAAAA
    AlaAspLysGlyProTrpCysPheThrThrAspProSerValArgTrpGluTyrCysAsnLeuLysLys

    435
1381 TGCTCAGGAACAGAAGCGAGTGTTGTAGCACCTCCGCCTGTTGTCCTGCTTCCAGATGTAGAGACTCCT
    CysSerGlyThrGluAlaSerValValAlaProProProValValLeuLeuProAspValGluThrPro

    458
1450 TCCGAAGAAGACTGTATGTTTGGGAATGGGAAAGGATACCGAGGCAAGAGGGCGACCACTGTTACTGGG
    SerGluGluAspCysMETPheGlyAsnGlyLysGlyTyrArgGlyLysArgAlaThrThrValThrGly

    481
1519 ACGCCATGCCAGGACTGGGCTGCCCAGGAGCCCCATAGACACAGCATTTTCACTCCAGAGACAAATCCA
    ThrProCysGlnAspTrpAlaAlaGlnGluProHisArgHisSerIlePheThrProGluThrAsnPro

    504
1588 CGGGCGGGTCTGGAAAAAAATTACTGCCGTAACCCTGATGGTGATGTAGGTGGTCCCTGGTGCTACACG
    ArgAlaGlyLeuGluLysAsnTyrCysArgAsnProAspGlyAspValGlyGlyProTrpCysTyrThr

    527
1657 ACAAATCCAAGAAAACTTTACGACTACTGTGATGTCCCTCAGTGTGCGGCCCCTTCATTTGATTGTGGG
    ThrAsnProArgLysLeuTyrAspTyrCysAspValProGlnCysAlaAlaProSerPheAspCysGly
```

FIG.1 CONT.

```
                                                                    550
1726 AAGCCTCAAGTGGAGCCGAAGAAATGTCCTGGAAGGGTTGTAGGGGGGTGTGTGGCCCACCCACATTCC
     LysProGlnValGluProLysLysCysProGlyArgValValGlyGlyCysValAlaHisProHisSer
                                                                      EcoRV
                                                                        |
                573
1795 TGGCCCTGGCAAGTCAGTCTTAGAACAAGGTTTGGAATGCACTTCTGTGGAGGCACCTTGATATCCCCA
     TrpProTrpGlnValSerLeuArgThrArgPheGlyMETHisPheCysGlyGlyThrLeuIleSerPro
                                                        StuI
                                                         |
                596              601                     |
1864 GAGTGGGTGTTGACTGCTGCCCACTGCTTGGAGAAGTCCCCAAGGCCTTCATCCTACAAGGTCATCCTG
     GluTrpValLeuThrAlaAlaHisCysLeuGluLysSerProArgProSerSerTyrLysValIleLeu
       ApaLI
                619
1933 GGTGCACACCAAGAAGTGAATCTCGAACCGCATGTTCAGGAAATAGAAGTGTCTAGGCTGTTCTTGGAG
     GlyAlaHisGlnGluValAsnLeuGluProHisValGlnGluIleGluValSerArgLeuPheLeuGlu

                642
2002 CCCACACGAAAAGATATTGCCTTGCTAAAGCTAAGCAGTCCTGCCGTCATCACTGACAAAGTAATCCCA
     ProThrArgLysAspIleAlaLeuLeuLysLeuSerSerProAlaValIleThrAspLysValIlePro

                665
2071 GCTTGTCTGCCATCCCCAAATTATGTGGTCGCTGACCGGACCGAATGTTTCATCACTGGCTGGGGAGAA
     AlaCysLeuProSerProAsnTyrValValAlaAspArgThrGluCysPheIleThrGlyTrpGlyGlu

                688
2140 ACCCAAGGTACTTTTGGAGCTGGCCTTCTCAAGGAAGCCCAGCTCCCTGTGATTGAGAATAAAGTGTGC
     ThrGlnGlyThrPheGlyAlaGlyLeuLeuLysGluAlaGlnLeuProValIleGluAsnLysValCys

                711
2209 AATCGCTATGAGTTTCTGAATGGAAGAGTCCAATCCACCGAACTCTGTGCTGGGCATTTGGCCGGAGGC
     AsnArgTyrGluPheLeuAsnGlyArgValGlnSerThrGluLeuCysAlaGlyHisLeuAlaGlyGly

                734
2278 ACTGACAGTTGCCAGGGTGACAGTGGAGGTCCTCTGGTTTGCTTCGAGAAGGACAAATACATTTTACAA
     ThrAspSerCysGlnGlyAspSerGlyGlyProLeuValCysPheGluLysAspLysTyrIleLeuGln
                                  ApaLI
                757               |
2347 GGAGTCACTTCTTGGGGTCTTGGCTGTGCACGCCCCAATAAGCCTGGTGTCTATGTTCGTGTTTCAAGG
     GlyValThrSerTrpGlyLeuGlyCysAlaArgProAsnLysProGlyValTyrValArgValSerArg

                780                                 791
2416 TTTGTTACTTGGATTGAGGGAGTGATGAGAAATAATTAATTGGACGGGAGACAGAGTGACGCACTGACT
     PheValThrTrpIleGluGlyValMETArgAsnAsn
                                                  SphI
                                                   |
2485 CACCTAGAGGCTGGAACGAGGGTAGGGATTTAGCATGCTGGAAATAACTGGCAGTAATCAAACGAAGAC

2554 ACTGTCCCCAGCTACCAGCTACGCCAAACCTCGGCATTTTTTTGTGTTATTTTCTGACTGCTGGATTCTG

2623 TAGTAAGGTGACATAGCTATGACATTTGTTAAAAATAAACTCTGTACTTAACTTTGA
```

FIG. 1 CONT.

```
GAATTCCGCA GACATTCCAC CCAAGACCAT TGGGCTCCCA CCTCTACTCT TTTGCCAGTT   60
AATGAATAGG CAGGAATTTC ACTGCCTGGA AAGAGGAACA ATGCTTTCTG GTCCTTATTT  120
CACATCTAAA ATAGAGAGGT CAATTGATTT ATTCCTAAAT ATCTTTGAAC ACTAAAATAG  180
AAGTTTTACA GCATATATAC TACCTGGTTG CTCTAGACTT AAGCCAGGGA AAAGTACAGA  240
TTCAACATTT AAAATTGAGA TAGACGCTTT CCACTTAATG CTACCAGTCT TGCTTTATTT  300
CATGAGAATG AGAATATAAT AATATGGCAT ACGTTCATTT GGGGGAAAGA TTGATGTCTT  360
ATAACATAAT TTATAATTAC AGAAAACATG TGAGTTCACT GGGAATAAAT AAATTTTGAA  420
GATAATAAGA TACTTTCACT TATGTCATAA TTTCTATGTC ATTTGGTGTA GGATGTAGAG  480
ATATTAACGT TTACACCTAA CTCAAGTTTG TCATCTAAGA CCTGAAAGGG TTTTGTCTAT  540
CAGCTGCACC CCTGGGTAGA GACACAACCT TGGGGAAGGC CTCAGCCCCA TCCCTCGTAC  600
AGCAGGAATG AGAACAGCCC TGCCTGTTGG GAAGCTTGAG GGAGGCTATG GACGTGCAGC  660
GCTTGGCAGA AGGTCTCGTC ATGGAAGGTT CCAGCAAATG TGAGATACTT TTATGATTTC  720
ATTTTCTCCA AAAGAAAGGG AATAAGAGAA GAGGGGAGGA AATAAGACTA ATTGCGAGAG  780
ATAAAGTACA AGGGTGAGGG AAGGAATAAG GAGACATGAC GGCAGCGTGG AGCAGCCGAG  840
GGGGGAGATT GCTTTCACCA CTTCCCAGCA TCTATTGCAG ATTCCACCCT CAAACATTTT  900
GTAAGGACTC TTTATTCAAG GTAACGTTTG AACCCTGCTG AGCCAGTGGC ATGGGTCTCT  960
GAGAGAATCA TTAACTTAAT TTGACTATCT GGTTTGTGGA TGCGTTTACT CTCATGTAAG 1020
                                    I
TCAACAACAT CCTGGGATTG GGACCCACTT TCTGGGCACT GCTGGCCAGT CCCAAAATGG 1080
                                          ↓
AACATAAGGA AGTGGTTCTT CTACTTCTTT TATTTCTGAA ATCAGGTAAG ACATAGTTTT 1140
TTTAAATTAT AATAATTATT TTTTCTCCCA CAATGTAGTA AAAATACATA TGCCATGGCT 1200
TTATGTGCAA TTCATTTAAT TTTTGATTCA TGAAACTTCC AGTTGAAAAT CTTGTATAAG 1260
ATTGAGGAAT TC                                                      1272
```

FIG.2A

FIG. 2B

```
CCCCAGTGTC TTTAGTTGCC ATCTTTATTT ATGTCCAAAT GCCCGACTGT GTGTTCTTAA   60
CTAAACATTT TGATTCATAG CTACCCATTC TACTTCCAGT AAACAGAAAG TTTTATTTGG  120
TTAATGCTAA CCAAATAGAT TAAAAGGAAG TCATGACAAT TAGACATTGA CATTGATTTA  180
CTGACCATTT ATTCCACTTG GATCTCCCAC CTCTAGGTCA AGGAGAGCCT CTGGATGACT  240
ATGTGAATAC CCAGGGGGCT TCACTGTTCA GTGTCACTAA GAAGCAGCTG GGAGCAGGAA  300
GTATAGAAGA ATGTGCAGCA AAATGTGAGG AGGACGAAGA ATTCACCTGC AGGTATTTCC  360
ATTGTCGTTG CACCTACGCA GGAATCTGTA ATTCAGATGG CAAGTAATTT ACTCACAAAT  420
TTATTAATGA TTTAAGAGGA AAGAGAAATT TATGGAGCCA GAGTTTGGAA CTATATTTGC  480
TCACAGTATG TGAAGCCATA CTAACAGCTT CTTGTTAAGG TTTATTGGAG TCTTTGTTAG  540
AAAAATACCC TCAAAGGAAG TTATTTGTTT TTACACCGGA CACAAACATT AGCAGTTATT  600
GTTCTGAGCT CCAGTTTTCA ACATCATCAT CAGTAAATGT TTGTTGAGGA TCAGGTGAAT  660
GAAAGTGTCC TAGATAGATC TGAGCAATGA CTTATAGCTA CAAGATCCAG TGCCTGCCCT  720
TTAGTATTTA AGGTGTAGTC AAAGAAACTG GATATAATGT TAAAAAAAAA AAAAAGACAG  780
CCCAAGTGAG GTACAGGCAT AATCAATGCA TGCTCTACCC AGATCCAGAA GAAAGAACAG  840
TGCCTAAGGT TGAGGCAGCT AGAGAAGGCT CAGGGAGGAG GTGGGAACTG AGCTGGGTTT  900
GGAGTTGAGA GAGCTCTTGA CAAGCACCAG GAAGGCAGGG GAAGATGCGG CCCTGCACCT  960
TCTGAGGGGG ACCATTAAGA GATGAAGTTG ACTAAAGCAG AGACTTTGTG TAGGTGACGG 1020
GCTTGGGAAG GTAGCTATGG AATCCAGACT GAGCACCCAT AGCAGGACCA CGGGATGGAG 1080
ATGGGAGGGG TCAGGGGCCA GGGTGGGGTG GAATGTGGAG CAGAGGTTCA GGGGAACTGA 1140
TCAGAGTTGG GAGGTCATGG AGACGGACTA TCTTGGCGAA TGGGTTCAAA GCAACCAGAG 1200
TTGCTTCTTT CCAACCCAAA AACAAAAATT AAGAAGATGA GTGAAGAAGA AGTAAAGCAG 1260
TTGAAACAGG AAGAAAGGGA AAATTATGAG GGAGGGAAGG TAAGGGCAGA TAAGATTTGC 1320
TGCCACGTTG GTGTATTTTG TTCAGTACTT CATCGATGCC ATGCCCAAAT AACTGAAAGA 1380
GGCAGCAATT CTGAACTCTC TGGTCCCTCA AGATATTCAA TGATCTTTAG CATGTCTCAC 1440
TTATTAATAA ACATTTGTTT TCTTTAAATA AAGAAAAATA CTTATTGGAT TTCCTGCTTC 1500
GTTCTGCAGG GCATTCCAAT ATCACAGTAA AGAGCAACAA TGTGTGATAA TGGCTGAAAA 1560
CAGGAAGTCC TCCATAATCA TTAGGATGAG AGATGTAGTT TTATTTGAAA AGAAAGGTGA 1620
```

```
GTACATTTTC TTCCTCCTCC TCCTACTGTC CTCCCCATCC TCCCACTCTT CCTCTTTCTC 1680

TATTCTATCT TTAATTTATA AGACCAGAGG AGGAAGGCAC TATCGTGTTA TAAAACTGAA 1740

TTC                                                               1743
```

FIG. 2B CONT.

```
CCAAGACCTC TGGCTGCACT GTGCCCCGTG GTGTCCCCAG CATCCTGGTG GGGCTCGATA   60
CACAGAGAGC TCATAAGTAG CATTTGAATA CATGAATCAA AGAATGGCTC AGTTTACTGC  120
AGCCTTTTTG CAGATGCAAA AGATGATCTT TTAGAAAGCA GAAACAGGGG GTCTGGTGCA  180
TGAGATCTTT TTCTCAACGT GACTATGCTG TGCAGACCTT CATGTGGTGT CTTGTGAAAG  240
ACTTTGACCA CTGTGTGGAC TTCCCTTCAG TGTATCTCTC AGAGTGCAAG ACTGGGAATG  300
GAAAGAATTA CAGAGGGACG ATGTCCAAAA CAAAAAATGG CATCACCTGT CAAAAATGGA  360
GTTCCACTTC TCCCCACAGA CCTAGGTAAG ACATTCCCTT TCATCTTTGT GTTCATCTAC  420
TGTAAAGTTG TCCCTCTGTG TCTGTGAGGG ATTGGTTCCA GGACCCCTGT GGCTACCAAA  480
ATCCATGCTT CTCAAGTCCC TTATATAAAA TGGTGCAGTA TTTGCATATA ACCTACATAC  540
CTTCTCTTGT ATAATCCCTA ATATAATGTA AATGCTATTT AATCGTTGTT ATACTGTATT  600
GTTTTTATTT GTATTATGTT TTATTGTCAT ATTGTTATTT TCTGTCATCT TTTTCAAGTC  660
TTTTCCATCC ACAGTTGGTT GAATTTGTGG ATCTGGAACC CCATGGATAC AGAGGGCCAA  720
CTGTATTTAG GATAATTTCA TCACTTTTAA TTCAAACCAC AATATGTGAA TAAGCAGATA  780
GAAAGAATCA AAAAGATGTC GATGTTCAAC TATTTTTGGC ACCATAGTAG AACATGGTTG  840
CTTTCTATTT TTTCTTGGAT ATGGAGGTTT CTTGAAGACC TAGAACATAG AAGAATGCCT  900
AGTTTAAAAA AAATCAATGA AACTATGAGT TTTAGGCCAA ATCTGAGAAA AGATCAAAGA  960
TGACTATGTT TGGGACTGAA GTAAGCATAT CAAGTTAGAA CTCTCATCAC ATGTTCGACT 1020
CAAATTGTGG AGCAAAAGAG TAAATAAGAT ATAAAAATGA AAATGAAGAT ACGTGAAATT 1080
CAAATGTTGC AACTTGCCTA TTATTTATTT TAGTGCATTT TTTTGTACTT TTCCCAGTTT 1140
GGTGTTAGGT GGCATTAAGT TCTCAGTAAT GACGCTTATC AAATAGGAAC TTAGTGCTTG 1200
TTACTCACCT TTATCCATTC CCCCAACACT CAACAAATTG CCTTTGCTAT ATCCCTATGA 1260
GATGAGCAGA TCAAATATTC CCCGTGAGTT AATGAAAACT GATTCAACCA AATGGCAAAG 1320
TCAGAGACTA TCGGGGGCCA TGGAGACACT CTGGGCCATT TTTATGAGGT AGTCTAGGCT 1380
CATCTTTATG AGGGAACTGA GGTCTCGGGG GGTGGGGG                          1418
```

IV

FIG. 2C

```
CCTCATAGCT ATTTACACTT AGGCAAGTTT TGTTTTGTTT TGTTTTACGT TGCCACTCAG   60
TTTTCTCATC TGTAAAATAG GGATAATAAC ACCTTCCTCA AATGGTTTTA TTAGGACTAA  120
AAGAGAGAAT GTGTGGAAAG ATGTTAGTGG AATTCCTGGC AGATAGTTCA CATGGACAAA  180
ATGGTATTAA CTACAAAAAT TTTTACAGAG AAAACGGTAA CTGACAAAAG CAGGTGTTTG  240
GAATGAATTA AGACCATGGC AGCCTTTTGA GGCCTTTATA TTTCTCCTGA CTGTGCAATA  300
AAAATATTTT GGCTCTCTAA GACTTGGCTG TCACAGTAGC AATGGTAATA TTAGCTACTG  360
TGCCAGAAGC AGCCTATCAA TAGAGAAATT GAAAATCTGA CCACACAAAT GCTGCAGCAC  420
CCAGCTGAAA TGCATTTGGA TGACAATCTC AGATGGGAAT CGAGAGCATC TCCTTCTGCC  480
TTGCTAATAG CAAGCTGATT TTTAGAATAT AGTCTAAGTG CTTCTTTTCC ATCCTCCCCA  540
GATTCTCACC TGCTACACAC CCCTCAGAGG GACTGGAGGA GAACTACTGC AGGAATCCAG  600
ACAACGATCC GCAGGGGCCC TGGTGCTATA CTACTGATCC AGAAAAGAGA TATGACTACT  660
GCGACATTCT TGAGTGTGAA GGTCAGGAGT GGTTCTAGAA AATGTTTTCA TTTCTGCCCT  720
TCACCTGTAA AATAATTTGT TGTAAAGCCC CTTCCCACAG GGATGTTATT AATAATTGAG  780
TAACGTATTC ACCTCTCGGA AAGAAGCAAA ACCCCAGAAT TAACCTGAAT TTTTTTTTTT  840
TTCTGAGACA GAGTTTTGCT CTCGTTGCCC AGGCTAGAGT GCAACCGTGC AATCTCGGCT  900
CACCACAACC TCCGCCTCCG GGTTCAAGAG ATTCTGCTAC CTCAGCCTCC CAAGTAGCTG  960
GGATTACAGG CATGTGCCAC CATGCCTGGC TAATTTTATA TTTTTAGTAG AGACAGGGTT 1020
TCTCCACGTA GGTCAGGCTG GTCTTGAACT CTCGACCTCA GGTGATCCGC CTGCCTCAGC 1080
CTCTCAAAGT GCTGGGATTA CAGGCATGAG CACCATGCCC AGCAGACCTG AATTATTTTT 1140
ATTAAAATGT TACATCAACA TGTACAAATA TAAAACTACA TCTAAACTCT AAGTACAAAC 1200
TTCTTATGCT TACAACTCTT ACACAGTG                                     1228
```

FIG. 2D

```
TTTTAAAAGA TCATTATTGA AATGAAGATG CCAAATATTG AAAACTCCTA ATGGAGAACG   60
TAGACTCCTG GGAATATATG CACCCTTGGC TCCCCACTGG CCTGTGCATC CCGGTCTAAG  120
GACATGGCAT CATGGAAATT CTGAACTTGG TCATGACTAC AATAGTTGAG GGAGTATTGA  180
CTAAAATATG TGAATGTTAC GGTTTAAAAG GAAAATGACA TTTGGATTAT GCTAGAAAAT  240
CCTGAGTCCT TATTGCCAAT TTTATTGCCA AGTGCCTGTT GTGAATTACA TCGGAATGAG  300
AGGCAAGTCG CACTTAAGTG AGTAGGATTC TGGTTTTTAC TCTCTATTTT GCTTCATCCA  360
TTTCAGTTTT CTTCTTCCTC TCTGTCCTTC CTTCCCACTC TGTCCAGAGG AATGTATGCA  420
TTGCAGTGGA GAAAACTATG ACGGCAAAAT TTCCAAGACC ATGTCTGGAC TGGAATGCCA  480
GGCCTGGGAC TCTCAGAGCC CACACGCTCA TGGATACATT CCTTCCAAGT AAGTCTCACT  540
GGGAAAAACA TTCCATGTTT AATTAAGGCT CTGCAGCTCT ATCAGACATT TGCTGTCATT  600
TAGATATTTT AGCATTCCTC AAGAAGTGAA CGCCTGATGT TTTTAATTTC AAAGCTAACC  660
TCCTCCCACA ATATTGCAAG TGAAATACGC ATTCTTGCTG CTCAAAATAT GGTCCACGGG  720
TCAGCAGCAG GGATGTTTTC TGAGAGTTTG TTAGAAATCC AGAA                   764
```

VI

FIG. 2E

FIG. 2F

```
CCAAAATGAT AAGGTCACTG ATTCTGTTGA GTGATTTTTA CACATGTAAA CTGTTAGAAA   60
AACAGTGCTT GGCAGCCGGG CATGGTGGCA CATGCTGTAG TCCTAGTTAC CTAAAGGGCT  120
GAAGCGGGAG GATTGCTTGA GTGAGTTCAA GGAGTTCAAG GCAAGCCTGG GCAATAAGTG  180
GGACCCTGTC TCTAAAAACA AACAAAAAAA AGAAAGTCCT TGGAATACAG GGCCAACCTT  240
GTTTCCTAGT TGCCATCTCT GAACACAGCC TTCATCTGAT TACCTCCTCC ATGCCCGACT  300
GTGCCTAGCA CACAGCAGGT GCTCAATGTT TGCTCTTGAA AAAGAGTCTT ATCCATGAAT  360
GTAAATGTTC AGTGCTACTA AAATCTTTCT TGTCCATTCA GATTTCCAAA CAAGAACCTG  420
AAGAAGAATT ACTGTCGTAA CCCCGATAGG GAGCTGCGGC CTTGGTGTTT CACCACCGAC  480
                                VII
CCCAACAAGC GCTGGGAACT TTGCGACATC CCCCGCTGCA GTGAGTATGA TGCACACCCA  540
GATTCCAGGA TTTGGACCTG CCCTGTTCTT GAAATCAAAA GAAAACATGT GTCAGTGCCT  600
GAGTGCAGCC TCTGAAAAGT GACCTACAAG TCCTATGGGA TGTTATTGGT CTTTATTTTA  660
TTGCTGGTTT AAAACAGTTA TGGTTATTGG TTACTGTGGG TGATTGATCA GAGCGTCCAT  720
TTATCATGTT TTTCTTTCTT TGCAACTGAA ACTTCTGCCT CAGGAGTTCA CTGAAATGTA  780
GGCTTTAGGT GTTGTTCATC CTATTCTCTC TGTGCTAAAG GGAAATCAGA CCCATGCTCT  840
CTGACACATG GATTTCATTT TCAACCAGAG TTCTAATAGT TGTTTTGTAA ACAAAGAGTG  900
TCTTTCTTTA CAATGTTCAG GTCTGTGGGT GTCCAGTTTT TCCACCTTGG GGAGCAGAGG  960
GTGAGTGGTG GGGGTGGGGA AGAGTTCAAG AGGAGAAGAT GAAATGGCAG ACCTAGTAGA 1020
AATGATGTGG AGTAAACAAT TTTATCATAT TTTCCTCTCT GAGAATTTGA AGCAAAGGAT 1080
TACACACTAA GAGAAATACA GGCATGAAAG GTTAAAAAGG ATTCAGTGAG GGTTGGCCTC 1140
CCCTCCTTTC CTCTGACATG TGTCCTTTGA AAGCGGAAGT TCCTCAGGCA TTCTCCCTTT 1200
TTATGAATAT TAATTTCTCT TTTTTTTCAG TTTCTCTTTT TGTCATCTTT TTTCCTCAAG 1260
AATATCTTGA TTTCTGGATG CACACACTTT TCCTTGGAGG TGTTTTTTGC CTTCTTTCCA 1320
TGGACTCTTT CCCTGTTGTT TGGCTTTTAT GGCATGTTGG GTGCCATTCA GTCATGTCTA 1380
CTCAGTGAAT AATTTATTCT TCAGGAAAGA GAGTGGACCT TTGGTGTATG TGAGAATTCG 1440
GGGTGTGAGG TGACACGTGT TGATACTTAC CAGGTAGGAA GAACTGAGCA AAGAGAACAT 1500
AGAAAGAAGC ACCTACCCAA GGGTCTTTCT CTGAAGGAGT TCCTTGTGAA AGGGTCTCAC 1560
AGGCATAGAT GCTACTAAAT TGATTTCATC TGAAAACATG AAACAATTCT CAAGTGCCAA 1620
```

```
ATTCCAAGAG AGGCTGAGCA GAAGCCAAGA CAGGCCAGAA CACCCTGCAG CCATCCTCCT 1680
TAACATCCAT CTGTGCATTC TCTATTTTAA AATTATTCAT TGTAGGGCTG GGCACGGTGG 1740
CTCACGCCTG TAATCCCAGC ACTTCCGGAG GCCGAGGTGG GTGGATCACG AGGTCAGGAG 1800
TTCAAGACCA ACCTGGCCAA TATGATGAAA CCCCACCTCT ACTAAAAATA CAAAAAAATT 1860
AGCCAGTTGT GGTGACACGC ACCTGTAGTC TGAGCTACTC GGGAGGCTGA GGCAGGAGAA 1920
TGACTTGAAC CCAGGAGGCA GAGGTTGCAG TGAGCTGAGA TCGTGCCACT GACTCCAGCC 1980
TGGGCGACAG AGCGAGACTC CGTCTCAAAA AATATATATA TTCATTGTAA CTTATTTTGC 2040
CCATTCAAGC AACACCTCCA CCATCTTCTG GTCCCACCTA CCAGTGTCTG AAGGGAACAG 2100
GTGAAAACTA TCGCGGGAAT GTGGCTGTTA CCGTGTCCGG GCACACCTGT CAGCACTGGA 2160
                                VIII
GTGCACAGAC CCCTCACACA CATAACAGGA CACCAGAAAA CTTTCCCTGC AAGTAAGTCC 2220
CCTCCAGTCT CATTCTGCTG CTATGGAATG TGAAATCCCA TTGACTTTGC CTTAGTTTTA 2280
GTTACTGTAG GAACGCAGGA TAAAGTATTC TGGAAGAAAA ACTGATCTAG TCATAAGTAA 2340
AGGAAATGAA CTTTAGCACG TTTTTTCCCG TAACGGTTGT TCTCAAAGCG TGGTTCCCTA 2400
GACTTTTTTC TTTTTGGAAA GCTAAACTCA CAATCACTTC TTTTTCAGAA ATTTGGATGA 2460
AAACTACTGC CGCAATCCTG ACGGAAAAAG GGCCCCATGG TGCCATACAA CCAACAGCCA 2520
                                IX
AGTGCGGTGG GAGTACTGTA AGATACCGTC CTGTGACTCC TCCCCAGTAT CCACGGAACA 2580
ATTGGCTCCC ACAGGTAAGC AAGGGTATGG GAGCTTACTG AGGGCCCAAG TTTTCTCCTT 2640
ATTTTTGTAT ACCAGTGGCA TCATCACAAT ATACAGTAGC TTTGTAAGTT TAATGCTATT 2700
GTGGTCAGAA AGCCTGCCCT TATGATTTCA GTTTTTTTAG ATTTGTTGAG GTTTGTTTTA 2760
TGGTTCAGAA TATAGCCATC TTGGTGAATG TTTCATGTGC TCTTGAAAAG AATGTGTCTT 2820
CTGCGGTTGT TGGGTGGGGT GTTCCCTCAA GGTCATTTAG GTGAAGTTGG TTGCTGGTGT 2880
TCTTCTGTAT CCTTACTGAT TGTCTGTCTC CTCCTTCATT GACTACTGTG GATGAATGGT 2940
GATGTGTCCA ACTTTAACTG TAAATTAGTC TATTTCTCTT TTAGATCGTA ACTCTTTTGT 3000
ATATTTTGAA GCTCTTTTGT TAGGCACATA TGTATTTAGG ATGGTTATGT CTTCTAGATG 3060
AAAGGACCCC TTTATCTTTA TGTAATGTTT CTTCTTATCT CTGGGAATAT TTCTTCTTCT 3120
GAAGTTCTGA ACTCTCTTTA TGGTGATATA AATACAGTCT CACAGCTCTA TTTTCACTAG 3180
TATTTGTGTG ATATATCTTT TAAATTTGTA TGATATATCT TTTAAATTTA TCTGAGCTTT 3240
TAAATTGAGA TGTTCAAACC ATTTGCATTC ATGCAATTGT TAATAGAGTT GAATTTACAT 3300
CTACCATCAA GTTAGTTATT TCTCTTTGTC CCATTTAAAC TTTGTTCCTT TTTTCATCTT 3360
```

FIG.2F CONT.

```
TTTCTGCCTT CATTTAGATT GAGTTTATCT CCACTACTCA CTTAGTAAAT TAATTTTTAA 3420
TGGTTTTAGT ATTTTCCACA ATGTTTATAA TATACATTTT TGACTTTTCA CATTCCACCT 3480
TCAAATGATA TCATTCTACT TGACATATGA ATCCTTACAT CATTGCAGTT CTACTTCCTC 3540
CCTCCCAAAA TGCTATACTA TTACTCTTTG TAATAGAAGC TTACTTCTAC TATGTCACAG 3600
ATCTCACAAT ACATTGACAC TATTTTTGCC CTAATAGTTG TGTTTTAAAG TGATCAAGAA 3660
TAAAACTATT TTAAATATTT TCTTTATTTA TTTATTTTAC CATTTCTGGT GCTTCTCATC 3720
TACTGGGGTA GATCTCAATT TCCATCTGGT GTCAGTTTCT TTCTGTGAAA AACAACTTTT 3780
AGCATTTTTT GTAGCACAGG TCTGCTACTG CTGAAGTCTT TCAGATTTTG AGTGTCTGAA 3840
AAAGTATTTT GCCTTCAGTT TTTAAAAGTA ATTTTGCTGA ACGTAGATAC TGGGTTGAGA 3900
GTTTCATTAC TTGCAACACT TTAATGATGA TGTTCCATTA TCTTCTGTTT TAAATAGTTT 3960
GACTAGTAAT CTGATCTTTG TTCCTATGTT TTCAATAGGT CATTTTTCTC TGACTACCTT 4020
TAAGATTTTC TCATCTTTGT TTTTCAACAG TTCGACTATG ATGTGTTTAT TATTAATTTC 4080
TTTGTGTTTA ATCTGCTTGA GGTATTCTGA GTTCCTAGAT TTGTAGATTG TTGATTTTTT 4140
TCTTTTCTCT TTTTTCTTTT CTTTTCTTTT TTTTTTTTTT TTTTTTGAGA TGGAGCCTCA 4200
CTCTGTCACC CAGGCTGGAG TGCAGTGGCG CAATCTCGGC TCACTGCAAA CTCCACCTCC 4260
CAGGTTCAAG TGATTCTCCT GCTTCAGCCT CCTGAGGAGC TGGGACTACA AGCATGTGCC 4320
ACCAGGCCCA GCTAATTTTT GTATTTTTGG TAGAGACAGA GTTTCGCCAT GTTGGCCAGA 4380
CTGGTCTCAA ACTTCTGACC TCAGACGGTC CATCACCTTG GCCTTCCAAA GTGCTGACAG 4440
TACAGGTGTG AGCAACCGTG CCCAGCCTAG ATTGTTGATT TTCATTGTCC TTGTAAAATT 4500
CATAGCCATT ATCTGTTCAA ACGTTTCTTT TTGCACTTTT CTCTCTCTGT ATTTTCCTTT 4560
TGGGACTCTA AGTACCACGT GTTTGGGATT CTAAGTACCC ACAACATTCA TGTTGTTTCA 4620
TAAATCTTGT AAGCTTGTTC TCTTTTTTTT TCAGTAACTC TTTTTCATTC TTTGTGTTGG 4680
TTTGGATAAG TTCTGGTAAC CTATTTCCAA GTTTATGGAT TATTTTTTCA GTTGTTTCTA 4740
GTCATCTCCT CAGCCCATTG AGAGAATTCT TCATCTCTGA TATTATGACT TTTTTTCTAG 4800
CATTTTCATG TTACTCTTTT CTATAGTTTC CATCTTTGCT GAAATTCTCT ACCTATCTAT 4860
GCATACTGTC CACCGTTACA ACAAGATCCT TTAACATACT AATGTAGGTA TCACACAATC 4920
CCAATCTGAT AGTTTCCAGA TGGCGTCTTC TCTAAGTCTG GCTCTCTGGA TTGCTTTATT 4980
ATTCAACAGT GGCTTTTTGT TCCCCCTTGG GTTTTTTGGT GTGTCTTATA ATTCTTTAAT 5040
CAAACACTAG ACATTATAAA TAGAAGAACA GTAGAGGTTA CAGTAAATAT TATTTATACT 5100
```

FIG. 2F CONT.

```
TTGAAATGGA CACCCTTGTC TTGCAAATAT ATATCGTGGA TAATTGAGTC AATGTAGTCA 5160
CTAGTTTAAC TGAATTGGGA TTTGTGATTG CTAGTTTTAC CTTAAGTGCA CCACAGATAT 5220
AAATTCCTCC AGTGATGTGC TGCTGCTATC TTTTACTTAG AGTGGGGCCT GGGGTGCTAA 5280
AGAGTTTTCT CCGTGTTCCT ATCCATTCCC AGATTTCAGC AGTCACTGCA TGCCTGCACT 5340
ACAGAGGAGA TATCTTCATA CACATAATCT AACCCCATTG ACACTCGGCT GTTTCTTGTT 5400
ACTGAATGCT CACTTTTTGG TGGACGTAGG AGAATACTTA TCTCCCTGGT CTACCTCCCT 5460
CTTAGGCCAG TTGAGCACAG CTCGGCTTTG AAAGTAGTGA TTTTTCAGTG TTCTTGTGCC 5520
TCCTTCTGAT GGAACTTGTA CCTGTGGTGG GTTTGGAAAG AAAGAGTAGT AGGCTTCTGC 5580
TTCATTGCAA TGCAGGATGT TGGGCACAAG AGGATTCCCT GTAACTTCTC CAAGGGAATA 5640
AGATTTTTGC CTCCACCACT CTCTGAGAAG CTGTGGATCT TTGCCTGCAG TCCTAGATGC 5700
AGGACCATCA CCTGCCCTAT CACCCAGAAG CTTTGGTCTT TGGCTTTGTT TGAGGAAGGA 5760
GCTAGAGAAA TGTGCAAAGC TTTCATGTCT GCCCCCCACT GACAGCCACT CACCACCCAC 5820
AGCCTGCACT GCCGAATGCA TCCTCCTCTC ATCTGCCCTC GTGTTCTCAT GAACACTCAG 5880
TAGGGACCCA TAAAAAAGAG CTTGCATGTA AGTGCAATTT CCAATTATAA GTACTCTATC 5940
TGTTCTTTCA CACCCAGGTT TTAAATGAAA TATTACTAGG AACTTATTAA TGTTCTAAAA 6000
TGCTATAAAT CTATTTTTAT GTTAATCTGT CTGCTAATAC AGAAAAGAGA ACAGTCATAA 6060
TTCTCAGAGG CTACCGTACT GTTTTTGTCA TAAATTGCTT CATGCTTCTT TTTTTTCAGT 6120
AATTGTTAAG CTTGATTTCT TTTATTTTAA TTTCAGCACC ACCTGAGCTA ACCCCTGTGG 6180
TCCAGGACTG CTACCATGGT GATGGACAGA GCTACCGAGG CACATCCTCC ACCACCACCA 6240
CAGGAAAGAA GTGTCAGTCT TGGTCATCTA TGACACCACA CCGGCACCAG AAGACCCCAG 6300
AAAACTACCC AAATGCGTAT GTCTTTGATT TTTACTGTAA GAGGGGCATC AGCCAACTGA 6360
AATTTCTGTT AAAAGAGCCA TGCTTCATGC TTCAAGCCAA CTTCCTAGGA CCAAATTTCT 6420
CTTAGACCCA GAATGTGTAG AAAAATGTCT CAAGAATCTT GCTTTTGAAG AAAGGGCCTG 6480
CGAGAAGAGA AATTTTAGGC TGGCTATTTT TCCTGAGTAG TTTTATGGAT GCAGGAGGAC 6540
ATCTGGAGGT GATGAGGTCA CATTAATTGA AAGCTCAGGA GTACATATGA GCAAATGCTT 6600
AGAAACAGTA CCATTCCACA ATGCCCACTA AATATCAGTG CAATATTTCT ACCATAGAAA 6660
TCTATCATTT TAACCTCCAA CCCCTGAAAT GAAGGTTGAA TTTGCTATTT TTGTCTTGGG 6720
TCACAAGTAA ATATACTTTA TATATATAAG TATGAATATA TATACACACA TATATATGTA 6780
TACATATGTG TGCATATATA AATACACACA TATATGAGAT ATACAAGTAT ACATATATAG 6840
```

FIG. 2F CONT.

TGTGTATATA TATGTACACA TATATGTGTG TATATATATG TACACATATA TGTGTGTATA 6900

TTAGAATATA TATAACATAA ATATGTATAT ATATATATTC TGACCTGTAT AAACACAGTG 6960

GATCCTGAGC ACCAGTGGCC TGAAAGGATA TGGGTTGCTG GGACATGAAG AACAAAAGCA 7020

GGATACGCAG ATGCTGAACA GCGAAAGAGG CCATTAGATG AACAGAAAAC CAGGTCTAAC 7080

AAGGACAGCT TTTCTTCCAT AAATGAGTAC ACAATATATG GAAAAAACTA TTTTTACATA 7140

TTGGAGAACA GATAAACTGA GATAATTTAG AAAGGGAATC AAATGAGATC AACCCAATAA 7200

CTACCTTGGC TTTGTTCCTG GAGACTTCCT GGGCTGAAGA ACAAGGAGAT GGAGCCCAAG 7260

CCGACCACAG CAGTCTTGCT GAACTGAGGA AGGAGACTGG AGTTGGGATT ACTAAAACAG 7320

CTGAGATTTT CTAGGCTAGG TAATAACATG AAAGGAAACA TTGTGGAGGA AAGCAGCTCC 7380

AGGAATGTCC ATAGAAAAGT CCTCAAGTCT TTGGCTAAAT AGAAAGCTGC ATATGCACAG 7440

GGAGAGGTTC CAGAGAGAAA ATAGGATAAA GAACAGCTAC TGGGGAAAGA AAAACTGCAG 7500

GGGAACAGTG AGCTCAATGG AGATGCCAGA GCTCACATAG CACTGGGGGA TATTTGAGTT 7560

CTGACCAGCC TGAGGAGAGA CCTCGCTGAA CATCTTGGGC ATTCAGTAGT CACCACATAA 7620

AGCCAAACTT TGGGAGTAGG ATTAGTGTAT TCCTATAATA AAGGCCACTC CAGAAACAGC 7680

ATAGTAAAGC TGAAAAGCAA GTCTAAAAAA ATCAACACGA TCTCCAAGTA AATTAACTGA 7740

TTGCCAGAAG AAAATTCAAC CCTTTAGAGG CAAACAACAA AATCAAGTTG CTCAGTTATG 7800

TGGCATCCAC AATGTGTGAC CTAAATTTAT AACTTTACCA GACATACAAA AAGCATTTAC 7860

TGTGATCCAT AACCAGGAGA AAAAGCACTC AAAACAAATA AACCCCAAAA TGAAGAAATT 7920

GGCAAGAAGA TTTGAAATAT ATATATATCA TAATTGTGTT CAAGGATTTA AATAAAACAT 7980

GAACATGGAA GAAACAAATG GATAATATCA AAAAAGAAAA ATTATAAAAT AACCAAATAG 8040

AAATTAAATA ACTAAAAAAG TGCATGTTTA ATGAAAAATG TACTGGCTAC CCTTACCATC 8100

AGGTTAGACA TTACAGAAGA AAAAGTTAAC TAGAAAATAA TTCAATAGAA GTGATACAAA 8160

CTGCAGCACA CACATACAAA GACTGAAAAG ATAAAGAAAC AGAGCCTCAA GAATATCTAT 8220

GAAAATATCA AAAGATTTCA TATATGTGTA AAGCAAGTCA CAAGAGAGGA AAGAGATATT 8280

GGGACAGAAA AAAATACTTG AAGCAACAAG AAAAATCTTA TTAGAAGCCA GAAGAAGAAA 8340

ATATATGTTT ACACAGAAGA ATAGTGGTAA AAATGACTGA TGCCTTCTCG TCAGAAACTA 8400

TGCTGGTCAG AAACAATGAA ATAACACCTT TAAAGTGATA GAAAAAAATA AAAAAGATTA 8460

ACATAGAATG TTATATCCAG CAAAAATATC CCTTGAAAGT GAATGTTATA TAAATACATA 8520

TTCTGCCTCC CCCAAAATAA ATAAAACACT AAGAGAATAT TTCATTACTA GGCTTATATA 8580

FIG. 2F CONT.

ATAAAAGATG TTCTAGAAAT CTATTTTGGT AGAAGAAAAA TAGTGCCAGA TGGGAACTTT 8640

ATACTAAGTA ATGAAGAACC CTGGAAATGG CAAATGTAAA AGATTCATAT TTAATGCCTT 8700

AATTTCTTTA AAAGATAATT GATGGGAGGC TGAGTCGGGC AGATCATGGG GTCAGGAGTT 8760

TGAGACCAGC CTGACCAACA TGGTGAAACC CCATCTCTAC TAAAAATACA AAAATTAGCT 8820

GGGCATGGTG GCACGTGCCT GTAATCCCAG CAACTCAGGA GGCTGAGGCA GGAGAATCAC 8880

TTGAACCCAG GAGGTGGAGG TTGCAGTGAG CTGAGATCGT GCCATTACGG TCCAGCCTGG 8940

GTGACAGAGC GAGACTCAAA ACAAACAAAC AAACAAACAA ACAAAAAGAT AATAATTTAC 9000

TACTTGAAGC AAAATGATAG CAATGTATTG CTACTTTAAC ATATGTAAAA GTAAAAATTT 9060

CTAAATAATA ATAATCACAT AAATAATGTA GGAAATAAAT GGTAGTATAC TGTTCTAAGT 9120

TTCTTGCATT ATCCATGAAG TTATATAATA CACATGGTTG AAGGTGGTAA GTTAAAGAGG 9180

GTTATTGCAA ATCCTAGAAC AACTGAAAAA ATTTAAACTT AGAGGAATAG ATAATAATAA 9240

GAATGTTCCA TTTATCCAAA AGAAGGAAAG AAAGGAAGAA AAAAGAATGA AGAAGATATG 9300

GCAAAGAGAG AAAATACACA GCATTATGGT ACACTTAAAC TGAACTGAAA ATATATTTAA 9360

TATACTCCTA AGCATATTAA ATATAAAGGG ATTAAACATT GCACAGAAAA GGCAGAGATT 9420

ATTAAGCTGA ATAAAAATCA AAGCCCAATT ATGTTCTTTT TACTATACAT GCTCTTTAAT 9480

TGTAAAGAGC TAGTCCAAAA ACCAAGTGTG GAAAATGACA TATCATGAAA ATAAGAATCA 9540

GAAGAAAGCT GGAGTGGTAA TGTTAATCCC AAAGTAATCT ACAAGAAATA ATACCACGAT 9600

GAAAAAGTTA TTTCTTAAGT AAAAAAAGTT TATTCATCAA GACTTAACAA TGCTAAATGG 9660

GTTGCACCCT CATAAGAGCC CTTCTGATAT ATGAAGCAAA CACTGACAGA ACTGAAGAGA 9720

CAAACAGATA AGCCCACAAT TAGAGTGGGA GATATCCTAA TGTCTCTCTC CGTATGGTTA 9780

TACATCTTCC CAAACAAAAT ATAATAGAAA AAATACACAA AAAAATCAGA AAGAATATAT 9840

ATGTTTTAAA GGAAATTGTC AACCTATTTA ACACTATGCC AAACTGCAGA ATACACATTC 9900

AAGTATGCAT GGAGCATTCC CCAACATATA CCATATGTGT GGGCCTACAG CAAGTCTTAA 9960

TAGATTGAAA AGAATTAAAA TGATACAGAG TCTGTTTTTG 10000

FIG. 2F CONT.

```
AGCAAAACAG AATTAAATGA GATATAAATA ACAAAAAAAT TGGGAAATTA TCAAATATCT   60
GAAAATGAAA CAACACATTT CCAAATACTT CATAAGTCAA AGAAGGAATT TAGAAAAGTT  120
TTGAACTGAA TAATAGTAAA AATACAACAT ATCAAAGTTC GTATGATGCA GCGAATGTTT  180
TTAGGGTTTT ATAACTTTAA ATGCTTTCAG TAGAAAATAG AAACATGTAA AAATCAATGA  240
CTTAAGATGG CATTTCTCAA AGTATGCTCT GGAGAAACCT GAAGTCTCTT GAGATCCCTT  300
CAGAGACAGT CTATGAGGTT AAAACACCTT TAAATTTAAA AAAAAAAAGA TTTTATTTGC  360
TATTTCACTT TTATTTCCTG ATAAGTGTAC AGTGGAGTTT TCCAGAGGCT ACATAATGTT  420
TGATCACATT ATCTCTCTGA TGGCTAATAA AATGTGTGAT TGTCTATTAT GTTTAAAAAC  480
ATTCTCAGTT TTGGATGCAA TAAATATTCA TAGTATATAT TACAAAATGA AAGCTCTTTA  540
GGGTCCCCAA TACTTTTTAA GAGTTAAAGG GTCTTAAGAC CAAAAACTTT GAGAACTGTT  600
GATTTAAGAT AACTTAAACA TCTAGAAAAG GAGAAGCAAA TAAGATCCAA GGTAAGTGGA  660
AGGAAGGAAA GAATGAAAAT CTGTGAAATC CAGTGTATAA GAATATAGAC AAACAATTGA  720
GTAAATCTGT GAAACAGAAA GTTGGTTCTT TTGAAAGATT CATGTAATTG ATAAACCTCT  780
GCCTAAACTG ACGACAAAGG AGGGAGCACC ACCGTCAACA TCAGGAGTAA AAAAAGGGAA  840
GAGTCATTGC TATAGGATCT TTTTGATATT AAAGCTAATA AACAAATATT GAGAGCAACT  900
TTACGTTAAC AAATTCAATA ACCTAGATAA TATGGACTAA TTCCTTAGAA AAAAACAAAT  960
AAGCAAATTG GACACTGAAT AAACTGAATT TCTAACCAAT CTGATATCTA TTAAAGACAA 1020
CATGTGTATA TAATCTTTAA TATGTTAATA TATATTAATA AATCAATAAA CTTCCCACAG 1080
AGAACACTCT AAGTTCAGAT GGCATCATTA GAAATTTTAT TATTTAAAAA AAATCCAATT 1140
CTTCACGATC TGTTACAGAA AATAGAGGAG AAGGGAAATA TTTCTTGACT CAATTTGTGA 1200
GAAAAAAAAA AAACCCTAGT TGTAAAAAAG TAGACAAGGA TATTGTGAGA AACTATAGCA 1260
CATTATGTAT TGTGAACATA AATATAAAAA GATGTAACAA AATTTTAATC ATTAACATGA 1320
TGAATATCCC AAACAAGTGA AGCTTCTCTT CAAGAATGCA AGGCTGGCTT AACATTTACA 1380
AAACAATCCA TGTAATCCAA CATGTTAACA GAATAAAAGT GATAAATCAT ATGATTATGT 1440
CAATAGATGC AGAAGAAAAT GTGACAAAAT TTAACACTTA TCCATGATAA AATGTCTTAG 1500
CAAACTATGA ATAGACTGGA ACTTCTTTAA CTTGATCAAA GGCATCTACA AAAGACCTCC 1560
AGATAACATC AACTTAATGG TGAAAGATTA ATGTTTTCTC TCTAAGATTG GGAATAAGAA 1620
```

FIG. 2G

```
AAATATGTTT GCTCTCAGTA CTTCTAATCA GCATTTTACT ACATTGGTCA CAACCATTGC 1680
CATAAGACCT GAAAACAAAA CAAAAAGAGA GGAAAAAAAG GAAGGAAAGA AAGAAAGGGC 1740
CTAAAGTTTG GAGAGGAAGA ATTAAAACTG CCTGTATTCA CAGAAAGCTT AATTAACGGA 1800
TGCAGAAAGT CCTAAAGATT AATAATTAAA TTTTGCAAGA TTGGAGAACA CATAAGTATA 1860
TACATGATCA ATATAATAAA AGTAGTTGTA TTTTTATACA CTGCCAATGA TCAACTGGAA 1920
AATAAAAATG TCAGAGCAAT ACCACTGACA ATAGTATCAA AACCACAAGA TATTTAGTGA 1980
TACATTTAAC ACAATATGCA CAAGAATTAT GTACTGCATA CTAAAAAACA TTGTTAAGGA 2040
AGGAATCAAA AGATCTAAAT AAAGATATAT CACGCTTATA TATTAAGAGT CAATATCACT 2100
TCTCACCAAA TTGATCTTTG GATTCAGCCC ATACCCAATT GTTAAGGAAG AAATTACAAG 2160
ATCTAAATAA AGATATATCA TGTTTATATA TTAAAAGAGT CAATATCACT TCTCACCAAA 2220
TTGATCTTTG GATTCAGCCC ATACCCAACC AGAATCTCAG CAGTCGTTTT TTTTAAAAAA 2280
TGTGAAAAAA TGTATATGCT AGAATCACAA GGACAATATT TAAAGAGAAG AAAAAAGTTG 2340
GAGGACTTAC TTACCCAAAG GTAAAGACCT ATAAAGGTAC AGTAAACAAG ATATGTGGTA 2400
TTGGGAAAAA AAAGTATACA GATATAGAAA TGGATGGTCC AGAAACAGAT CCACATATAC 2460
ATGATCAATT TAGTTTCTAG GTAGGTGACA AGGAAATTCA ACAGGGAAAA ACATCTTTTC 2520
CAAAATCATT GTGAAACAAT CGGATATCCA TCTAGAAAAC AAAAATAAAA ACAAATTTTG 2580
ACTTCTACTT TCCATCCCAA ATTAATGTGC AAAAGCTCCT AGATCTAAAT GTAAGAGCTA 2640
AAACTTAAGC TGAAATAAAA CAATTCCAGG AAAATATATA ATATTTTCAC AAACTTGAGG 2700
AAGGCAAAAT TTTTTTCAGG CAGGACCCAG AAAACACTAG CTTTAAAAGA AAATAAATTA 2760
TAATTTGGGC TTTCATAAAA TGAAAATTAT GTTCATCAAA AGTCATTGTT AAGAAATCAG 2820
TAGGTAAGTA ACAGACTGGA ATAAAAATTC TCTCCATCCA TATATCTGAC AAATGGTTTG 2880
TATCTAGAGT ATAAACGTTT CTCCCACTCA CTAATCAGAG GACAAACACC TAATTAAAAT 2940
GGGCAACAGA ATTGAATAGG AAATTTCTCA GGGAACGATG GACAGATGGA CAATAAGCAC 3000
CTGAAAAAAA TGCTCAACAT TTTAGCCATC AAAGATATAA GAATTATAAC CATCACAAGA 3060
TGTCACCAAC ACTTAATTGG CATGGGTATC ATTAAGAAGA CACAACAATA AGTGCTGTCA 3120
CTGATGTGGA GCGAGGATGT GCAGCTCTCG CATACGCTGG TTAAAGTACA GTATGCTGGT 3180
TTTCCATAAA GTTAAATAAC TATGAGTCTA CCCCAAAAAA CTGCAATTCT ATTCCTGAAT 3240
ATTTACCCCA TGGAAATGAA AACAGAAGTC CACAAAGAGA TCTACAAGAA TATTCACAGC 3300
AGCTCTAGTT ATTATAACCC CAAACTGTAA ACAACTACAA GGTCAATCAA TGAGAAAATG 3360
```

FIG. 2G CONT.

```
AATCGATAAT TTGTGATCTA TTCATATAAT GGAATATTAT TAAGCAATTA AAATGAAGAA 3420
GTGACTGATC CTCTCAAATA GGATGGATGG AACTCAAAAA TATATTAAGG AAAGGAGGCA 3480
GATACATAAG TGTACATTCT GTATGAGCCC ATTTATATCA GGTTTGAGGA GAGGTAAAAC 3540
TAATCTTTAG TGAAGGAAAC CAATAGTATT TTCCCTCTGG CAGTGGGAAG AGGGTAGCAG 3600
GAATTGAATG AGCAGTGACA CAGGGTGTTT CTAGAGTAAT GGAAGTGTTC TGTATCATAT 3660
GGGAGTGTGG TTTACACAAG TATAGGTGAT CATCAAAACT CACCAAACAA CATTTAAGAT 3720
CTGTGCATTT CACACTATGT AAAAGTATAC CTCAACTGAA GAGAGTGGAA ATCTGTTTCA 3780
AATGCTCAGC CTTTTAACAC ATCCAGTTGC TTAGACTATG AACTTCCTCA AATGGGGTGT 3840
CTGGGCTTGA GATTAGATCA CATGTGTAGA GTCGCTAGAG AGACAATGTT GCATTCCCAT 3900
GGTACATAAT ACATTTCCCG TTTTCTCAGA CAGCCACAGG TCATGAATGT GAGGATTCTG 3960
AGAGGTTGGA GCAACATTCT TGGGAGGCAT GAGGGGGAGC ACATTCTCCA AGATCCCCCC 4020
CAGCCCGGGG TCCTCGCCTG CTTTGACTAT TACTCCGTTG TTTTCGGACT CCTCCGTAGC 4080
TGCCCGACCT CTTCAGATCC CATAGTCTCC CTTTATATCT TGAGTCCCAC TGTTCTTCCA 4140
ACTCATCCCC CATTCCCTCA GACCTGGAGT GCAGTGGCCA GCAGAGGATG GATTGAGAGC 4200
AGGAGAGGAT GTCCTGCCCA GGAACCCATC CTAGAGAAAT GCATCCTGCC TGGGAGCTAG 4260
TTTCCCAGGG TGGCTTTGAT ACGTCTTGCA GAAACAAACC CACTTGACAC ACCTGATACG 4320
GTATTGACAG TAACACTATT TTTCGTGGTT GTTTTTCATA GTAAAAGTAG ATCCCTTTAG 4380
TTACACTGTG AGTACTTAGA GTAAGGTGAC TGGCCTGGGA ATGATACCAT CTTGGATGTC 4440
ATTTTCTCCT TGGAGAAATG TATTTTAGTT CCAATGCACA TTTCACAATA CAGTCCTATA 4500
GAGAGAAATA CAGAGAGCTA GACAGTTAGA GATATACTTT TATGTGCATA AAAATATAAA 4560
ATATGCACTT TAAAATCTGT ACCTGTTATT CCTGAGAAAT GTATTTGGCA GAAGGTGGGA 4620
GGGGGATATT CTGATCCTTT TATTTACATG TTTATGTATG ATCTGAGTTT TTATATGGAG 4680
CATATACTAC TTTTGATTTT TTAAAGAAAA ATTAAAATCT GTCTTTGAAA TGTACACAGT 4740
TGTTTAGAAG TTGAGGACCA TTTTTGTTTG TTACAACATT ATTGTACCTA TAATGGGAAT 4800
ATTTCAAAGC CACTTGTTAA CACTTTGTTA GAACAAAATG TAGAGGGTGC TGGGTGCCCC 4860
TGAATATTCT CCCACCTCTT GTGACCTGTA TTGTTTTGGA ATTTCCAGTG GCCTGACAAT 4920
GAACTACTGC AGGAATCCAG ATGCCGATAA AGGCCCCTGG TGTTTTACCA CAGACCCCAG 4980
CGTCAGGTGG GAGTACTGCA ACCTGAAAAA ATGCTCAGGA ACAGAAGCGA GTGTTGTAGC 5040
ACCTCCGCCT GTTGTCCTGC TTCCAAATGT AGAGACTCCT TCCGAAGAAG GTAAGAAATC 5100
```

XI

FIG. 2G CONT.

```
TGTGGCTGGA CATCTACACG CTTGGACGCT GGGATGAAAA GCCATGGAAA ATCTCACTGA 5160
TGCAGAAACC TTCCATGCTA CACGAGAAAT CAAGTGTTTT TAGAGGGTCT GCCATGTGGA 5220
AGGAAGCCTC AGTGCACTCT CTCAAGGAGG CAGAGGTGTG ACTTTTGGCA CAACGTGAGT 5280
GGGCTGTGCC TTTAGGACAG GTGCAAACCC TCCAAGGTGC TCAACTTAAC CACTCACCTT 5340
GTTCTAAAAT GGGTTATCTC AGTATCCCAG TCCAAATTCG TATTCTATCA TGCTGCCATA 5400
TGTGTGATTC TTTCCAAGCC AGTAAGCATC TCCAGTAATT TCTTAAGGTA GGCAGCGTTC 5460
ATTGCAGTCT TCAGCATTGC AGTTTCTGAG GAATGTGGCC CCTGATTCTG TCATCCTAGA 5520
GAAACCTGAC ATGACTGTAT TGATTCCATA TCATCCTGGG TCTCTGTGGC TCTTCATAAT 5580
CATCCATTTT TTCCCTGTAC AGACTGTATG TTTGGGAATG GGAAAGGATA CCGAGGCAAG 5640
AGGGCGACCA CTGTTACTGG GACGCCATGC CAGGACTGGG CTGCCCAGGA GCCCCATAGA 5700
                                XII
CACAGCATTT TCACTCCAGA GACAAATCCA CGGGCGGGTC TGGAAAAAAA TGTAAGCCAC 5760
TTTGATTTGG ACTCTTTGGC CTTTTGCTCA CCAATCTTTG CAAACAGAAT TGGTTCTGTG 5820
TTACAGAAAA TCTGACCTGG ACTGCTCTTT TTTGTAATGG GGGAGAGGGG ACAGAAGAAA 5880
ATATTGGAAA GGCATCAGGG GGCTACGCTA GAATATAATT GGCCTTAGTA TGGAAAGTAC 5940
AAGCAGCACA GGCCAGGAAA CCTCCACACA TGTGAGGGTT CTCAGGCCTC TTCCCTTTAG 6000
TGACATTTCT TTAAAGTTTC CATTATTGGG GACTGTCTCT AGTTTCTAGT GTTTGTATGC 6060
TAGGTTCCAG TAATCAAAGA TGCCCTTTAT GAAATTTAAG TCAGATTTTT CGAGAAAAAA 6120
TTTGGATGGG CCATCAGGTC ACCATGGGAC TTCCCTTAGC CTCATCGATT CTCTGCGATG 6180
GTTTACTTTG GGGCCTATGA ATAGGGAAGA CTGAGATATA GGAAAAACCA AAGTGTCTGT 6240
GTTCCCCCAC TCTCACACCC ATGCAGCATA ACACTTCTCA CACCAGATGT GGGGGGATTT 6300
CTCCTCACAC CCCAAGCGAG TCTCCAGCAG ATACCAGCTG GTGTCCTACA ACGTAACGTC 6360
AGTGCTGACA CTCTATCTGG AGACAGCGTC AGATCCCATA AGTTAAGGCT CAGTCCCACA 6420
AGACCGCCCC ACTGCAGATG CCAATCCCAA GTTCCAGGCG GTGACCTGTA CTTCTGCCCA 6480
ACTGGACAAA AATCTGTTTT TCTACTTGAT TACTTTGCTA GAGTGGCTCA CAGAACTCAG 6540
GGGAACACGT TACTTTTATT TACCCATTTG TTATAAAAGA TATTACAAAG GATCCTGGTG 6600
AACAGCCAGA CAGAAGAGAT GCACGGGGCA AGGCATGTGA GAAGGGGCTC AGAGTTTCCA 6660
TGCCCTCTCC AGTGCACCAG CCCCCGGTAC CCCAAGTGTT CAGCAACCCA GAAGCTCTCC 6720
AAGTGCAGTC TTGCTGGGTT TTTATGGAGG CTTCATTACA GAGGCACAGT TGAATACATC 6780
GTTGGCCATT GGAGACCAGC TCACCTTCAG CTCCTGTTCC CTCCCTGGAA GTTGGACGTG 6840
```

FIG. 2G CONT.

GGGGGCTGAA CAGTTCCAAC CCTGCAATCA CATGGTTGGT TCCTTTGGCA ACCAGCCCCA 6900

TCCTGAGACT ATCCAAGAAC CCACCAAGAG TTGCTTCATT CAAACAAAAG ATGCTCCCTT 6960

CACTCAGGAA CCCCCAAGGG ATTTAGGAGC TCCGTGTCAG GAACTGGGGG GCAGAGACCA 7020

AATATACGTT TCTTATTCTA CCACAGTGTC ATATGAATGG GAGGACAACA CTGCCTTTCT 7080

GTGTCTTGCC CCATAGAGGG CGCACAATGC ATGGAAATAA ATGTTTCTGA ATCAACAGCA 7140

AACAGGCTTC ATCGGGTAGG AGAGCGCTGA GCCCTCCAGG GACAATGCAC ATCAATGATG 7200

TCCCACTGTC CTTTGGTGCT GGGGCTCTAA GGCCTCCACT GGGTCAGGCT CCTGAAGGGA 7260

GACCCATTCT CCAAAGACCC CCGAGGGTCA CCACTCCCTG TCCAGGGGTG TGGCCTCATA 7320

GCTCCTTTTG AACAGGGGCA CAGGAAGGAC GGCTTTAGAG CATTCAAAAA ATAACTTTGC 7380

CAAAATAATA ATAATAATAA TAGAAAGAAA GGAAGAAGAG GCTGAGCATG GTGGCTCACA 7440

CCTGTAATCC CTACACTTTG GGAGGCTGAG ACAAGCAGAT CACCTGAGGT CAGGAGTTCG 7500

AGACTAGCCT GGCCAAAATG GTGAAACCTC ATCTCTACTG AAAATACAAA AAAAAATTAG 7560

CCAGGTGTGG TGGCGTGCAC CTGCAGTTGC AGCTACTCAG GAGGCTGAAG CAGGAGAATC 7620

GCTTGAACCC AGGAGATGGA GGTTGCAGTG AGCTGAGATC ATGCCACTGC ACTCCAGCCT 7680

GGGCGACAAG AGCAAAACTC CACCTCAAAA AAAAAAAAAA AAAAAAAAAA AAAGAAGGAA 7740

GGAAAAAGAA ACACTCCTTT ATGTCTTCTA AGGATAGACA TGAAATGCGT GAGCCTTGGA 7800

ACACCTTCTC CCTCTCCTGC CCCACGTGAG CTGGAGCTTA CATGCCTTCT TGTTTTCAGT 7860

ACTGCCGTAA CCCTGATGGT GATGTAGGTG GTCCCTGGTG CTACACGACA AATCCAAGAA 7920

XIII

AACTTTACGA CTACTGTGAT GTCCCTCAGT GTGGTAGGTT GCCTTCTTTT TGGTAAGGAA 7980

ACTGCTTACT TAATATGGAT TTGCAACAAA AAAGGAAAAG GGCTTCTGAG CAGACTGCTT 8040

CTGGGGAGGA GATAGCTGCC CTCTCCATCA GACCCCACTC TTCATCATGG GCATCTTGAA 8100

TCTGCCCTAC TATTGGCCAC ATTTGTTAGA GGAACACCTG CCCATCGCCC CAGGCACACA 8160

TAAATAAAAT AAATGTAAAA TTCCCAAAGA GCAAGCTTAG AGGTAATCTA GTCAGCCCCA 8220

GGATGGTCCC ACTGAATGCT GCCATGTCTA GCGTGGGATG CATGAAAAAT TTAGAGTCAT 8280

TCGGATGAAA AACTTTCCCT TTCCACAGCT GAGAAGTAAG AAAGAAAATA CAAACAGCAG 8340

GAAACAGGTA AGCATGTAAC GCACATTGTA AACCTCAGAT GGCCATCCTA GGAATTCAAT 8400

GAAAGGTAGT GCAGCTCTTT AGCCCCAGAT GGCCTTTCTT ATAAGTTTAC TACTCACAAG 8460

TCACATTAGT GACATAGCTT AGAGACTGCT TGTTGGGTTC CATCCTCATT GCTCTGAGAC 8520

TCTTGTTGGG AGTATGAGGC TTGGATCAGG GGAAGGGGAG TTGACATTAG TTCTTAAAGA 8580

FIG. 2G CONT.

```
ATTGGAATAA CAAATCCATG GGTATTTCTG AAAAAAAAAA AAAAAAAAGA AAGGAAGCTA 8640
CTTGGAATTG TCCCATATTT AACATTCTGC TGACCAATCA ATTTGTCCTA GTTACAGAAA 8700
ACCACCCTGG ACTTCTCCTA TGCATAATTT GGTTGCTTGT GGTTGGGTCT GCCATGTGGA 8760
GGGACCTTGA GCTGGGGGAA GGAGCTTGGC CTCCAAGTCC ACTGAAGACC AGCATCCTGA 8820
GATTGCCTGG GAAGGTGGTA CAGGGCAGTG ATGAAGATCA TGGGAGCCAC ACTGCCCAGC 8880
TTCGCATTTG GGCTTCTCCT AGGGACACCA AGAGGGAGGA AGGAGGGGTT AGGATGGTAT 8940
GAAAGATTCT ACTTGGCCAA TATTATTGTA ATGCGGCATT GTGATCTCTG GATTTAGCAT 9000
GAGTTGATAG CTGACTTTTT CTGCAGAAGC ATCTTGGTGG CACCTCTAAC TCAAAGTCCC 9060
TCGATGGAGT CAGTTCCAGT TCTCCACTTC TGGCCCCATC TGGTACACAC CACTGCCTCT 9120
CACTGCCTGG GCTCTCTATC CTTGACAGGC TGCCTTGAAG TTGAGCCCAG ACTGATTTTC 9180
TTGCCTCAGA CCCCACTACC GTGCCTGGGA CTCATGCACC TTTGACTCCC ATGGAAGGGA 9240
AGTGCAGTAG TTTCCCAGGT GCAATTCTGG TGTCCTCACC CACATTGAGG ATGTACAAGA 9300
ATCAGGTTCT TAGAGATTGG AGAAAGAAGG AAGAATGGGA ACAAGATTTT TCCCAAAGGA 9360
CTGTGAGGTC CCCCACCTAA CCTTGATGTG AGACAAGTGA GGTTAACCCC AAGCCTGGTG 9420
AGAAGCGTTC CCATCAGACA CTTGGAAATC CTGAGGACTG TTTCATGCAG AAGGATATGG 9480
TTTATTCAGG TTTGACTCAT GCTTGAGAAA GCTAGAGCCT CTGGTGGTGA ATGATTTTAA 9540
TAACTATTTC CTTTCCACCA ACATATACAG TACAAATAAT AATAAGCAAA AATAAATAGA 9600
AACATTCAGT TTTGTTTTGA ATAGTAGGAG CAGGGTACCA TCATTTCTGT AGTTACTCTT 9660
TTAGTACAAC GATGCATGTC TACTGTATGT AAGGCATACT AGCAGAAATT GAGCTCAGCA 9720
CTAGAGAAGA TGATTGCATT CTATGCCTTG CTTCTTTTTT AAAAAAAAGG CTTCCATAGA 9780
TAGATTCTCA GAACAGCCCA TGGCAAATGT AAAGTTATTT GGAAAACCCA GGTTCCAGAT 9840
TCACTAGAGC ATAGAATCTC TGGTTGGTTG GGAAGGAATT TCCTCTTACA GTTGTTACTA 9900
ATAATTGTAT GAACAATTAT TTAAAATATT AACATTTACA TTTGTGAAGA CCTTGAAGGG 9960
CTGGAGACAA CAGAGAAGCA TTTTTGAATA CCCTCTGCAG                       10000
```

FIG. 2G CONT.

CCCCTGCACT GTTGTAGGCA TTGGTGGATG GTACCAAAGA TGGGACACTG TCCCTACCTC 60

CAGAGACCCT GTGGGCTGGC TACAGAGAGA AGGCAGGGAG GAGGAAAAGA AGAATAAAGT 120

CATATGTTTA AGTCACCCCC ACGGCCGTTG GTTAGTCATG GGAGGCTCCC CAGAGGAGCT 180

GTCCTGAAGC TGGCTGACAG AAGGCAACAT TTCAACTTAG GACAGTAATC CTTGCTACAT 240

ACAATCACAT ACACACACAC ACACACGTGC ACACACAGAG ACTCACATGG AAAAATAAAC 300

CTTTGTGCCT TTCAGCAGTG ATGACAATTA TGGTTTTCAG TAAACTTTAC ATGGTTTAGA 360

TGGTGATGGT GATGATGATG ATTATGGGAA GGATGGCATC ATGTTCTAAA CATACTGCAT 420

GGAGTCAGAA TAACAATGAC AAATAACCAT TTGTCCCAAT CAAGGTTTTC TCAGAAAATA 480

TCTCATTCTG ATGCTAAACT ATACCAGTCT GTTTGATCAC TTCTCCAACA AAATAATTAC 540

AAAGTGCTTA TATTTTCTTG AAAAGAGAGG GTCCTGTGTT GTCTACTACC ACTTTTGAAA 600

CTTAGAGAAA ATGTTCCAAA AGATGATGAT TTTACTATTT AGTTCGGCCT TTAAGATGTC 660

AAAAACTCAG TGCTTGGAAT TTGTCTCGAA TTACACCACA AAATTGCTAC CTTGTCTCAA 720

ATGGGATTTC TTTCCCACCT TGTGCCACAG ↓CGGCCCCTTC ATTTGATTGT GGGAAGCCTC 780

AAGTGGAGCC GAAGAAATGT CCTGGAAGGG TTGTAGGGGG GTGTGTGGCC CACCCACATT 840

XIV

CCTGGCCCTG GCAAGTCAGT CTTAGAACAA ↓GGTAAGAACA GGCCCAGAAA CGATTTATAC 900

TGTCCCTCCA CGTAAGCCCT GCAAAACCCT TCTACATTTA CATAAAATCC ACACAGCTGA 960

GGCATCAGCA CCTGCCTCTA AGTTTTCTGA AGGAGGAAAA AAGCTACAAA AATTAATATA 1020

TGTATATATA CATATATATT TTTATAGGTT CTCTACTGTG AAAATGACAA AAATTGCTGT 1080

CTTTTTCTTG ATCTGGGCAG CTCCATCAAA ATCTGTAGGC ACAGTGATTT GCACCAAGTT 1140

CCAATATTGC TGGAAAATAC TGAAGATGCT CTGAGGATTT CTATGGATAT CCATTGTCTC 1200

ATTGTCAGAT GAAAAGAGGG GGAAGTTTTT AGAAATGTGA CACTTTCTGG GTTGGGAGAG 1260

CAAGGACAAA ATTATCTCCA GTCTATCACA GGCACAGATT CTTTTTCTTT GGACACTTTC 1320

GTGAATCATT GAATTCAATG CAGAGGCTAC TCATCCATTC GCAAACAAAA AAATTCTAGG 1380

TCATGATCCC CATAAATGAA GAGTGATCAG TCCAATCCCA GGGAACCTGG ACATTTTGGG 1440

TATTGTTTCA GTGGAACATG CCTTTCATAA GTTCCATTTT CTTGGGTATC TCTTAGGAAG 1500

CAAGCATAGG AAACAGGCCC ATCCGTCTGC CTGTTTTGCT TCCTCATCTC ACTTCTACAC 1560

GAGGGCGCCT GTGCTCAATT GCTGTTTTCC CCTAAAGAGA CTCTTTTCCA TAAGTTTGTG 1620

FIG. 2H

AAATGCCATC GACAAACCTG ATCGCATTGC ATTTCACTCT GCTGTTGAGT CGATTTTTCT 1680

TTATTTTATC ATTTAGTAAC TCCTTGCTCT ACAGAGCTTT CACCTTCCAC ATATTTCAGA 1740

TTCATTCTTT CCTAAACTAT GTGGTGGTCT ACGTCCTCAC TGACTTATCA ACATGCTACC 1800

ATCATGCACT TCCTATCTCT ATTCCTCTTC ATTAAAATCT GGTTCCAAGT GGCTCACACC 1860

ATTATTCTGA GCTATTACCT GCCTACGCAG TCCTAGAAAG TAAGTGATTC AGGAAACATT 1920

CCCCAAAAGT AAAGTTTCTC AGGTAAGATC AGAAGACTCC CATGAGTCAC TGCTGCTCAG 1980

GATCACATCT GGCTCCTTGA AGAGTGATTC ATCAGACCTT ACATAGATCT TGTCATAAAA 2040

ATGAAAGAGG CCTCGGGGGA AGGTCTTGGG CTGGTGGCTT CTGTTGGAGT CCTGGGCTGT 2100

GGGGTGAAAG CCGTGGCTGT AGAGCTTCAT GCGGAGTTAC TTAGCTTTGC TCTCCTGTGG 2160

ACAGGCCATG CTGTGCCTCC CCCAAGCATC GGAAAAATTG GCATAGATGG GCCCTTCTCA 2220

AAAATCCCAC TCCTGGAGCA CTGGCCAAAA TTACTACCAT CCTGATGCTG GGCTTGCAGT 2280

CCTTTCCTTT GGGAATATGA ACATGGTCAA AATTAAGTGA ACGTGTCTTT CTGGCTTTCT 2340

GTACAATGGA GCAGAACAAA GTATCAATTT AACTAAAATT TGAACTAAAT CCTCTTTCCA 2400

GGTTTGGAAT GCACTTCTGT GGAGGCACCT TGATATCCCC AGAGTGGGTG TTGACTGCTG 2460

XV

CCCACTGCTT GGAGAAGTAT GTTTAGGGGA CAATTGACAT GAAGTCTTGT CTTAAATACT 2520

TTTTCTGTCC TTCTTTTCCT CCTTTCCTCC TTTCCTTTCT CACTCTTCCT CCCTTCCTTC 2580

TCTGGCTGTG ACACTAGGGA CCAGGCCAGG GCAATTGGAT AAGAGAGAAG GGAAGGGTTT 2640

CTAGAAAGAA ACTGCAGAGG AAAGACACAG TACAGATGAT TTTGTGGGCC TGAATAAACT 2700

GCAGAACAGA GCTGTTCACT ACCATAGGCT GTATCAGTCT CTGCCCAAAC AGCCCAAGAA 2760

CATTCCTTAA CTGCCTGTTT CAAGCAAATC ATGAATTTTG CTTCTTGCCA CTCAGAAGTC 2820

ACTAATTCTG AGTGGCCAAG GGTGTCAGGG AGACAGCACC AATTTCATGG CACAGAGGTT 2880

ACCTGAAGGG GCTGGACCAT ATTTTCCTCT TGACATCCTC ATCTTTTCTA GGTCCCCAAG 2940

GCCTTCATCC TACAAGGTCA TCCTGGGTGC ACACCAAGAA GTGAATCTCG AACCGCATGT 3000

XVI

TCAGGAAATA GAAGTGTCTA GGCTGTTCTT GGAGCCCACA CGAAAAGATA TTGCCTTGCT 3060

AAAGCTAAGC AGGTACTCGT TCACCTGTGG TCTTCACCCC ACGCTGGTGA AGATATTTGC 3120

TTTATGTCTG GGTTTTATGG GCCATGGCAC TGCATGGCAG TGGGAGGAAC TGTCTATCAC 3180

ATGAAAGGCT CAAGGGCTTT GGGGACAGCA TCAATCTTCA ACCCTAGCCC TGCCACATGC 3240

TAGCTGTGCT CTTGAGAAAG GCAGCAGGAC TCCGTTTTCT CATGTGGAAA AAGAGTTGAA 3300

ATGAGGTACT CTGTTACTCC TAGAACTCAC TTAATGTTCA CCAGTTCATA CACATTCATG 3360

FIG. 2H CONT.

```
ATCAGAGAAC GATTCAGTTA TTCCAGGCTG ACAATTCCCC CTTCATCATA ATATGTTTAA 3420
GAGAATCATA TAAGACTATA TTTGTTTCAA AGCACTTTAA AAACCACAAG ATCGAGTTGG 3480
GTGTCTGGTG TGGGTGCCTG TAATCCCAGC TACTTGGGAG GCTGAGGCAG GAGGGTCACT 3540
TGAGTCCCGG AGTTTGAGGC TGCAGTGAGT TATGATCGTG TCACTGCATT CCAGCCTGGG 3600
CGACAGAGTA AGACACTGTA CCAAAAAAAA AAACACCAAA AAAACAAAAA ACAAACAAAA 3660
AAAAAACAAC TTCACAATGT CAAAAAAATC ACAAATACAG TTTATAAATG TAAATTATAT 3720
TATTATTATT GTCTTCTTTG ATTTGATTTT CTCTTTCCTG TTGAAATGTT GTTTCACTAA 3780
GCCTGACAAA GTGAAACATT TGCTTATGTC ACTCATTTAG TGCTGTTTGG AGCCAGATAC 3840
TAGTTGAGTC AGCTAAGAAA CAGCTATTTG TAGGAGAAGC AGGTTTGGGA CAGGTGACAA 3900
GGCACGCAGG GCGCTCGCTG TGCTGGTGGT TCTGGAAGAC AGGGTGTCAG TGTGGACAGG 3960
GATGAGCATG GCCTGGATGA GAAGGCACGG GGCAGGAGCC TGAGCTGCTC TCCTGGGCCT 4020
GGCCACAAGC CCAGGGCAGC TTCTCTGGGT CTGTGAACTG AGGGGTGATG TCCTGGGATG 4080
CTCTGACACT CTAGAAGGAG AGAAGAGCCT TTCCAGCTCA GCCTTTATAA ACAGTAGCTG 4140
ATCTCCCTCC TGCTCCCCAG TGTCCTCCCC GCCATCCCAG CAAATGTGCA AATAGAAGGT 4200
CCCCGTTCCT CATGATCCTC AGAGAGCTGG GGTGTTCTGA TGGCTTGAAC AAGTAATTTG 4260
GAAATTTTGG GTTTTGGAGG AGTTCTCTGA TAGGCTGATA CATTTCGAGT TTAGAGTTCC 4320
CACCCCACAT CCCCACACCC CGAGTCTAGG GCATTTAGTG CTCCACCAGG GAACCTGTAG 4380
AGTGAGGAAG TCTGCATGAC AGGCTGGGCC TTCTGATGAT GCTCAGAAGC AGAAAGTGTG 4440
CCTGCTTCAA AGTTGGTGAC GATGATGTTT CTTGATCAGA ATAGGGCATT TCTTATTTCC 4500
AATCCTTTAT CCTCTTGAAC TTACTAAAGT AGAATCAGGT CTAAAAACCG GAGTTCTAAT 4560
GTTTGAGAGT CCCTGGGACT CTAAAGTATA TGAATGTTCT TTGAAAACAA ATACCATTTT 4620
GTTCAAGCAA AAGGCTTATT TCCAATCCTC TTTCATTTGG TATCAAGTAT TTTACTGGAT 4680
TCTTACAACT ATGGCGTAGT AACATTCACT GAGGAGGAAA TGGAGGATCC AAGGATGGAG 4740
CAAGTTGCTC TGGGCACACA ACACATTTGC AATTTTACAG CCTCTTGGTG GCATCTCAGT 4800
CAGACATTCC ATGCACTGAT CAATGCCCTA TTCGATTAAT GTAAAAGGAC ACACTCAGCA 4860
TGAGATTCCA GTTGTGCACA GAATACTACA TGAGAAGTGC GCCTTTGTCA TCCCTACTTT 4920
CAAAGGTGAA GGCCACCAGC AGTATCTTGC ATGCAACTGA TGCCTTTCAA ATGAAACCTT 4980
ACATCTGCAT AGTCCATAGA CAACCACAGG CAAATGTGAG GGTGAAACTC TGTGTTCTAC 5040
GTTGCTCTGT GTCAGTGAAG CAAGGCAGTG CAGTTCAGAG GGCTCTGGGG CCTCAAGACA 5100
```

FIG. 2H CONT.

```
GGGATGACTG GTTGTGGGTA CTGCAGCTGC GAGCAGAGCA GTCAAACATA ACTGCTGATG 5160
CTTTTCTTTC AGTCCTGCCG TCATCACTGA CAAAGTAATC CCAGCTTGTC TGCCATCCCC 5220
                                  XVII
AAATTATGTG GTCGCTGACC GGACCGAATG TTTCATCACT GGCTGGGGAG AAACCCAAGG 5280
TGAGATAAAT TCCATTGCCC ACATAACGAA TTGGTTTTGA CCTACAGTCC ATGTGACAAA 5340
ATGATCATTT TGGAGAAAGC TGTGCAAATT CCTATCCATG AATGTGGTCC ACCCCACTCC 5400
TGATTTTGCC TGGGCACCTG TCTATGTCTT AATCAGTCTT CAAGGCACAT GATCAAAGGG 5460
AGGAAAACTG TGTCTTTGAG TCTCTCTCTC TCTCTCTGTT TTCAGAACAT TTTTATTTCA 5520
ATTAATTAAT TTTTAACTTT TATTTTAGGT TCAGGGGTAC ATGTGCAAGT TTCTTGTATA 5580
TGTAAACAGT GGTTTGTCAT GCAGATTATT TTGTCACCTA GGTACTAACC CTAGTACCCA 5640
ATTCTTAGTA TTTCCTGCTC CTCTCCCTCC TCCCACTCTT CTCCCTCAAG TAGGCCCCAG 5700
TGTCTGTTGC TCTCTTCTTT GTGTCCATGA GTTCTCATCA CTTAGCTCCC ACTTATAACT 5760
GTGAACATGT GGTATTTGGT TTTCTGTTCC TGTGTTAGTT TTCTAAGAAT AACGGCCTCC 5820
AGCTCCATTC ATGTTCCTGT AAAAGATATT ACCTCATTCT TTCTTATGGC TAAACAGTAT 5880
TCCATGGTGT ATATGTACCA CATTTTCTTC ATCCAATGTG TCATTGATGG TCATATAGGT 5940
GATTCCATGT CTTTGCTACT GTGAATAGTG CTGCAATGAA CATTCATGTG CATGTGTCTT 6000
TAGGGTAGAA TGATTTATAT TCCTCTAGGT ATATCGCCAG TAGTAGGATT GCTGGGTTGA 6060
AAGTTAGTTC TGCTTTTAGC TCTTTGAGAA TCACCATACT GCTTTCTACA GTGGATGAAC 6120
TAATTTACAG TCCCACCAGC TGTTAGTGTT CTCTTTTCTC TGCAACCTTG CCAGCATCTG 6180
TTATTTTTTG ACTTTTTAGG AAGCCATTCT GGCTGGTGTG AGATGATTTT TCATTGTGGT 6240
TTTGATTTGC ATTTCTCTAA CGATCAGTGA TATTGAGCTT TTTTTCATAT GTTTGTTGGC 6300
CACAGGCATG TCTTCTTTAG AAAAGTGTGT TAGTGTCCCC TGTCCATTTT TTAATGGGGT 6360
TTTTTTTTTC TTGTAAATTT GTTTAAGTTC CTCATAGATG CTGGATATTA GACCTTTTTC 6420
AGGTGCATAG TTTGCAAATA TTTTCTCCTG TTCTCTAGGT TTTCCCTTTA CTCCCTTGAG 6480
AGTTTCTTTT TCTGTCCAGA AGCTCTTAAG TTTAATTAGA TCCCATTTGT CAATTTTTGC 6540
CTTTGTTGAG ATTGCTTTTG GCATCTTCAT GAAATTTTTG CCCGTTCCTA TGTCCAGGAT 6600
GGTGTTACCT AGGTTGTCTT CCAGGATTTT TGTACTTTTG GATTTTACAT TTAAGTCTTT 6660
AATCCATCTT GAGTTGATTT CTGTATATGG TGTAAGGAAA GGGGTCCAGT TTCCATCTTC 6720
TACATATGGC TAGCCAGTTA CCCCAGCACC ATTTATTGAA TAGGGAGTTA TTTTCCCATT 6780
GGCTTGTTTT TGTCAGCTTT GTTAAAAATC AGATGTCTGT AGGTGTGTGG CCTTATTTCT 6840
```

FIG. 2H CONT.

GGGCTCTCTA TTCTGTTCCA CTGGTCTACG TGTCTTTTTT TTTTTTTTTT TACCAGTACC 6900

ATGCTGTTTT TGTTACTGTA GCCCTGAAGT ATAGTTTGAA GCCAGGTAAT GTGATGTCTC 6960

CAGCTTTGTT CTTTTTGTTT AGGATTGCCT TGGCTATTCT GGCTCCTTTT TGGTTATATA 7020

TAAATTTTTG AAGTAGTTTT TTAATAGTGC TGTGAAGAAT ATCATTGGCA GTTTGATAGG 7080

AATAGCAATG AATCTGTAAA TTACTTTGGG CAGTATGGCC ATTTTAATGA TATTGATTCT 7140

TCCAATCCAT GAGCATGGGA TGTTTTTCCA TTCATTTGTG TCATCTCTGA TTTCTTTGAG 7200

CAGTGTTTTG TAATTCTTAT TGTAGAGATC TTTACCTCTC TGGTTAGCTG TATTCTTACA 7260

TATTTTATTC TTTTTGTGGC ATTTGTGAAT GGGACTGTGT TCCTGATTTG CCTCTGGGCT 7320

TGGCTGTTGT TGGTGTAAAG GGATGCTAGT GATTTTTGTA CATTGATTTT ATATCCTGAA 7380

ACTTTGCTGG AGTTGATTAT CAGCTGAAGG AGCTTTTGGG CTGAGACTAT GGGGTTTTCT 7440

AGACATAGAG TCATGTCATC TGCCAACAGG GATCGTTTGA TTTCCTCTCT TCCTATCTGG 7500

ATGCCCTTTA TTTCTTTCTC TTGCCTGATT GCTCTGACCA GGGCTTCCAA TACTATGTTG 7560

AATAGGAGTG GTGAAAGAGG GCATCCTTAT CTTGTGCCAG TTTTCAAGGG GAATGCTTCC 7620

AGCTTTTGCC CATTTAGTAT GATGTTGGCT GTGGACTTGT CATAGCTGTC TCTTATTATT 7680

TTGAGATATA TTCCTTCAGT ACCTAGTTTA TTGAGAGTTT TCAATATAAA GGATGGTAAA 7740

TTTTATCAAA ATCCTTTTCT GCATCTATTG AGATAATCAT GTGGGTTTTC TCTTTAGTTA 7800

TATTTATGTG ATGAATCACA TTTATTGATT TATGTATGTT GAACCAAGCT TACATTCTGG 7860

GGATAAAGCC TACTTGATCA CGATGGATTG GCTTTTTTAT GTGCTGCTGG ATTTGGTTTG 7920

CAAGTATTTT GTAAAGGATT TTTGCATCAG TGTTCATCAA GGATATTGGC CTGAAGTTTT 7980

TTGTTGTTTT TGTGTCTCTG CCAGGTTTTG GTATCAGGAT GATGCTGACC TCATAGAATG 8040

AATTGGAGAG GAGACCCTCC TCCTCAGTTT TTTTGAACGG TTTCAGTAGG AATGGTCATA 8100

GCTCTTCTTT GTACATCTGG TGGAATTCAG CTGTGAATCT ATCTGGTCCT GGGCTTTTGT 8160

TGGTTAGTAG GCTATTTATT ACTGATTCAA TTTTGGAGCT CATTATTGTT CTGTTCAGGG 8220

AATCAATTTC TTCCTGGTTC AGTCTTGGGA GGGTGTATGT GTCCAGGAAT TTATCCATCT 8280

CTTTTAGGTT TTCTAGTTTG TGTGCATGGA GCTGTTTGTA GTAGTTTCTG ATGGTTATTT 8340

TTATTTTTGT GGCATCAGTG CTAACATCCC CTTTGTCATT TCTAATTGTG TTTATTTTGG 8400

TCTTATCTTC CTTTTCTTCA TTAGCCTAGC TAGCAGCCTA CCTATCTTAT TACTGTTTTC 8460

AAAAAACCAA CTACTGGACT TGTTGATCTT TTGAATGAAT TTTCATGTCT TGACTTTCTT 8520

CAGTTCAGCT CTGATTTTGG TTATTTCTTG CCATCTGCTA GCTTTGGGGT TGATTTGCTC 8580

FIG. 2H CONT.

TTGTTTCTCT AATTTTTTCC ATTGTGATGT TAGGTTCTTA ATTTGAGATC TTTCTTCTTG 8640

ATGCTAGCAT TTGGTGCTAT GAATTTCTCT CTTAACACTA CCTTAGCTCT GTCCAAGAGA 8700

TTCTGGTATG TTGTATCTTT ATTCTCATTA GTTCAAAGAA CTTCCTGATT TCTGCCATAA 8760

TTTCATTATT CACCCAAAAG TCATTCAGGA GCATGTTGTT TGATTTCCAT GTAATTGTAC 8820

GGTTTTGAGT TATTTTCTTA GTCTTGACTG GTATTTCATT GTGCTGTGGT CTGAGAGTGT 8880

GTTTGGTATG ATTTTGGTTC TTTGGCACTT GCTGAAGATT GTTTTATGTC CAATTATGTG 8940

GTTGATTTTT AGAGTATGTG CCACATGGTG ATGAAAATGT ACATTCAGTT GTTTTGGGAA 9000

AGAGAGTTGT GTAGAGGTCT ATCAGATCCA TTTGGTCCAA TGCTGAGTTC AGGTCCTGAA 9060

TATCTTTGTT AATTTTGTGC CTCGATGATC TGTCTAATAC TGTCAGTGGA GTACTGAAGT 9120

CTCCCACTAT TATTTTGTGG GCGTCTAAGT CTCTTTGTAG GTCTCTAAGA ACTTTATGAA 9180

GCTGGGTGCT CTTGTGTTGG GTTCACATGT ATTTAGGATA GTAGATCTTC TTTTTGAATT 9240

GAACCCTTTA CCCCTTTACC GTTATGTAAT GCCCTTCTTT GTCTTTTTTG GTCTTTGTTG 9300

GTTTAAAGTC TGTTTTGTCT GAAATTAGGA TGGCAACCCT TGCTTTTTTG TCTGATTTCC 9360

ATTTGCTTGG TAGGTTCTCC TCCATCCCTT TATTCTGAGC CTATGGGTGT CATTACATGT 9420

GAGATGGGTC TCTTGAAGGT AGCATACCAG TGGGTCTTGC TTTTTATCCA GCTTGCCACT 9480

CTGTGCCTCT TAAGTTGGGC ATTTAGCCCA TTTACATTCA AGGTTAGTAT TGCTATGTGT 9540

GAATTTGATG CCCTCATTGT GTTGTTATGC TGGCTTGTTT GTGTGATGGT TTTATAGTGT 9600

CATTGGTCTG CGTATTTAAG TATATTTTTG TATTGGCTGG TAGCCATCTT GCTATAGTTA 9660

GTGCTTCTTT CAAGATCTCT TGTAAGGCAG TTCTGGTGGT AACCAACTCC CTCAACATTT 9720

GCTTAGCTGA AAATGATCTT ATTTCTCTGT TGCTTAGGAA GCTTAGTTTG GCTGGATATG 9780

AAATTCTTGG GTGGATATTT TTTAAGAATA TTGAATATAG GCCCCAATAT CTTCTAGCTT 9840

GTACGGGTTC AGTTGAGAGG TATGCTGTTA GATTGATGGG GTTCCCTTTG TAGACGACCT 9900

GTCCTTTCTC TCTAGCTGCC TTTAACATTC TGTCTTTCAT TTTGACCTTG GAAAATCTGA 9960

TGATTATGTG TCTTGAGGAT GATCTTCTTG TATAGAATCT 10000

FIG. 2H CONT

```
CACAGGGGTT CTCTGTATTT TCTAAATTTG ACTATTGGCC TCTCTAGCAA GGTTGAAGAA   60
GTTTTCATGG ACAATATCCT GAAATGTTTT CTAAATTGTT TACTTTCTCC CCATCCCTTT  120
CAGAAATGCC AGTGATTTGT AGATTTGGCC TTTTTACATA ATCCCATGTT TCTTGGAGGC  180
TTTGTTCATT CCTTTTCATT CTTTTTTCTT AATTTTTGTC AACTGTCTTA TTTCAGAAAG  240
CCAGTCTTCC ATTTCTGAGA TTCTTTCCTC AGCTTGGTTT ATTTTGCTAT TAATACTTGG  300
ATTGCTTTGT GAAATTCTTA CAGTTTGTTT CTCAGCTCTC AGCTCTGTCA GATCCATTAG  360
GTTCTTTTTT AAACCAGTGA TTTTGTCTTT CAGCTTCTAT ATCATTTTAT TGTGATCCTC  420
AATTTCCTTG GATTGGATTT TGCCATCCTC CTGGATCTTG ATGATCTTCA TTCCTATCCA  480
TAGTCTGAAT TCCAGTTCTA TCATTTCAGC CAGCTCAGCC TTGTTAAGAA CCCTTGTTAG  540
AGAACTAGTG TGGTTGTTTG GAGGACATAT GGCACTCCGG CCTTTATGTT CCTTTAACTG  600
CAGTGTAGGT TGAATACAGC CAATAGACTT GTTCTTTGGA TGTTTTTACA GGGCCAAAGC  660
CTTGTGCAGG GTCTTTATTT GTAGTTGATT TCTTGTCTTT GGTTTCATAG TGTGGTATGT  720
TAGCAAGGTA TTTTTGGTGT TGAAGCTTTG GGGTGTGATC CATTTTTTAT TTGTATATTT  780
CCCTACACCT AAAACAAGCA AAAAAACAGT AAAGGTCTTT GAGTCTCTTA ATCCATAATT  840
TCAGCATTCC TGAGTATGCT TCCCTGGGTA AGTGGGGTTT TCACCCAGCC CTCAAGTTAA  900
GAGTGTTAGA TTATTTTTCA TGTGAAATTA GCCAGACTGG CTTTCTTAAC ACAATGTAAA  960
ACAATAACAA CAAAAGTTAT AATTAGACTA GTCTTCTTCC CAAATACCCA CATGTCTAAT 1020
GTAAGTGGGA TGGTGTTAAA CAGGGGACCT ACAACTGGGG GAGAGGCGGA CAGGTCCCAT 1080
GGCCCCAGGT CTAGGATGGC ATTTGGTATT GGTTGATGGG TGTGGATGTG AACAAGAGAG 1140
GGAACACTTG TGCAGGATAT GGTATCAGCA CCTGTAATAC ATTTTAGGGA TTCTTTCTTC 1200
TCTTTGCAGT ATGCCCTGAC AATAATTATA TCCATCAGCC TAGTCCCCTT GGCCATTGAA 1260
ACACTAAGAC TGTCTTAGGA TCCCTGCTGC AGTTTCTCAG AGGTGCTAGG AGGGCATTAG 1320
GAGTCTGAAG CCCTGGAAGT GTGTTCTGAC TTTGCCACTA GCTAGATAGA CCTGGACTAG 1380
GCACGTTACC TCTTTGTACC ACTCAGCTCT AACCCCTCAT TCAAAAACCC AGCATTTTCA 1440
AGTGGTGTTT TTCACATCAG CCTTTGCATA AGTTTTCATT TGAAGAAAGG TTTTTTTGTT 1500
TTTGTTTTCT TGGTTTAATC AAACATTTAA AAACGAATGG TCTAGATGAT TTCAAAGTGG 1560
CTTTCCTTTT CCTGTGCTTT TCCTACTATT TAAAAACTTC ACCTCCTTGA TTTCTTGATC 1620
```

FIG. 2I

TCCCTTTCTG CACTGCTGGG TCTGGGAGCA TTGAGGCCAA GTAAAAGGAA CCTTGGCAAA 1680

GGAGGAACAC CTATGGGTGT GCCAGGCTGC TCCCAGTGTT TTGCATTTTT AAAAATTTAA 1740

ATGCTGCAAA CCTCTATGAA TTACATATTA TTGTTCCTAG TTTACAAATT AGGAGCCTGA 1800

GGCTCAGAGA ATGTGTGGGA TGGTACAGAC TAACCTGAAT TAGAACCCTG GCTCCCATTT 1860

ACTGGCTGTC AGGACTTAGA AAAGTCATAA ACTCTCTGGC TGGGTGCAGT GGCTCACGCC 1920

TGTAATCCCA GCACTTTGGG AGGCCGAGGC AGGCAGACCA CGAGGTCAGG AGCTTGAGAC 1980

GAGCCTGACC AACACGGTGA AACCCCGTCT CTACTAAAAA TACAAAAATT AGCCGGGTGT 2040

GGTAGCACAC CCCTGTAATC CCAGCTACTC AGGAGGCTGA GGCAGGAGAA TCGCTTCAAC 2100

CTGGGAGGTG GAGGTTGCAG TGAGCCAAGA TTGTGCCCAC TGCACTCCAG CCTGGGTGAC 2160

AGAGTGAGAC TCTATGTGAG AAAGAAAGAA AGAAGGAAAG AAGGAAAGAA GGAAGAAAAG 2220

AAAGAGAAAG AAAGAAAGAA AGAAAGAAAG AAANNNNNNN NNNNNGAAAG AAAGGGAAAG 2280

AAAGAGAACG AAAGAAAGAA GGGAGGGAGG GAGGGAGGGA GGGAGGGAGG GAGGGAGGAA 2340

GGGTGGGTGG GTTGTGAACT CTTGTTGATT GTTTCCTCAG CTGAAATGTG GGCTGCAGGG 2400

CTATTGGGGG AGAAACAATA AGAAAGTGCA CCAAGCACCA AGCACATGCT AAGAAGTCCA 2460

TCATGGCAGC TCCTGATAAT AATATGGAAT AGAGTTGTAT CTAACATGAC TCTTTCTTGC 2520

AAGTGACAGA AAATGCAACT TAAGTTGGAT TAAGCAAAAA AGAGAAATCA TTAGTGAACT 2580

GAAAATTCTG CAGGCTCACA TCATGGCCCC AGACCCTGTC CATTATTCTT GGGCACAAAT 2640

GTGACATTCT CGTGGCTGCA GATGCTGTGG TGGCTCTGGC TCTGCAGGAA AAGAAATAAG 2700

GAAGGCCACT CTCCCCATTA CACAAACAAC AGTCTTCCAG CTCTGAGAGG TCGAACTTGT 2760

GTCACCAGCC TGCCCCTAAA CCCGTCACTG ATTAACTCCA ACCTGCATCA GCTGTTCCAT 2820

GCTGGAGGTG GACGCAGGAC CACACTCATA CCAAGATGGG GGCAAAGTGT AGTTCCCTCA 2880

ACAGGATTAT AGGATATAGT GTGATAGGCT GCTGGGCCAG AAAAGCAAAC AGATCCTCTA 2940

CAATTCCTCA ACTGATGAAA GCACGAAGCT AAAATCATAA AGATCTGTGT GTGAGTTCTG 3000

GCTCTCCCAT CTTCCTTGTG AGATTGAGCA GTTAGTTAAT CTCTTTTAGC CTCAGCTTTC 3060

TCACCTGTAC CAACATATAA GGTCATTGTG AGGATTAAGA TTATGCCTCA TGATCATCAT 3120

TATCATCATC ACCATCCACA TTGCAACCAC AACTACCATC ATCATCCCCA CCAACATCAT 3180

CACCACCACC ACCATCACAA TTATCATTAC CACCACCACC ATTGTCACCC TCAACATCAC 3240

CATCATCACT ATCACCACCA CCATCATCAT CACTACCACT ACCAACACCA TCACTCTCAT 3300

CATTCCACCA CCATCACCAT TAACATTACC ATCACTATCA TCACCACCAC CACCACCACC 3360

FIG. 2I CONT.

ACCCCCATCA TTACTGCCAT CAACATCACC ATCACCATCA TCACCACCAT CACCATCATT 3420

ATCAACCATC ATCACCACCA TTCCACCACC ATCACCATTA TCATCACTAC CATTATTCCA 3480

CCACCATCAT CATCCACCAC CACTACCACC ACCATCACCA CCATCATCAC CATAACCATC 3540

ATCACCACTA TCAACATGAT AGTAATTATG ATTACCACCA CCATTAGCAT TATCATTACC 3600

ACCACCAGTA CCATCACCAT CACCACCGCC ACCACCTCCA TGATCATTAC TACCCACCAC 3660

CATCACCGTC ACCATCATTT CACTACCAGC ACAATTATCA TTACCACCAC CATCACTACC 3720

ACCCTTATCA CAACCCTCAT CATCACCACC ATTCACCAGT GCACCACCAC CACCACCATC 3780

ACTATCATTA ACAATAGACA TCACATAACC AGTTTGTAGC TGGACCTTGA GCCCAGAGCC 3840

CACTCACTGT TTCTTCAGTC CCACCGCCAA CCACCAGGAT GAGTCACAAA ACATAACTCA 3900

GGCCTGCTCC TCAATTTTCT ACATGTCAAT AATGACATTG AAGCAATGGG TGTTCTCTGC 3960

TTCTCAGAGG GAAGTTGAAA TTCTCCTGCT CTTCCCTTCA TGTTTCCAGA TGTTCCCTGA 4020

CTTGGATATT CCAAACGCAG AGTTTGGAGG TGTTGAGGCC AAGGGGTTTT TCCAGGTCAG 4080

CCATCATCTG CAATCACTGA GCTGATCCTG CTGCTGGACT TTCCCTGTTG CCCTCTCCCC 4140

AACGCCCATC GGGGAGGGCT TCAATCCTCA GGTCACCTGT GGCCTTTCTG CCCTCAGAGG 4200

TGCCATCTCT ACATCTACCA CTGGAAGGCA GCACCTACTC ACAGATTGCA TCAATTTCCC 4260

AGCAACTCAT GGTGGGTTTT CCCCCTTATC AGCGTGTTTG CCTTGCTCAG AGAGCAGATC 4320

CCAGAGCAGT GACACCTAAC TTAATTTTCA GCAAAACATT TTGAGAAGGG TGCTCCCTCA 4380

CACAACTACA CAGTCCAGGT GATGCACCCA CTGCCCAATG CTTGGTAGTC AAGAGGAGCT 4440

TCCTCCCTGC AGCTCTGCCC AGATAGGGCT GAG 4473

FIG. 2I CONT.

```
ATTGGGAGCT GCCTCGTGTT CTGCAGCCTC ACAGACAGGA GGTCCAGTGC CGCTGCTCTG   60
TTCTGGAATA TCCTCCTGAA TGTGTTTTGG GTGCAGTTGC CGTTTCTTTC ATCTTTTTAA  120
ACACAGGTAC TTTTGGAGCT GGCCTTCTCA AGGAAGCCCA GCTCCCTGTG ATTGAGAATA  180
                                  XVIII
AAGTGTGCAA TCGCTATGAG TTTCTGAATG GAAGAGTCCA ATCCACCGAA CTCTGTGCTG  240
GGCATTTGGC CGGAGGCACT GACAGTTGCC AGGTAAGCAA AGATCAAGAG ACCAAAGTTA  300
GTCTTGTGCT CTCTTGTCTC AGTCTCAGCC CCTCAGACTT CATTCCCCAG GTGGCAAATT  360
CAAGGATTTT CAACCGAAGA CCCCAGTCTA AGTGTTGTTT AGAAACTTCC TAGATCTGTC  420
CCTGAATGCG TATTCAGATC ATCTAAGGGG ATGTCTTGGG GCTTGAGTTC CAAATCAGTA  480
GCAAGCGAGT TTTAAGTGCC ATAACTACCT CAGGCCACTC ACCCTCCTGG GGTGTGCTGG  540
TGGCCAGGGA CTAAAGTGGT GACTTTTCCG GTAGGGAAGG AGGTAGAGGG TACAGGACAG  600
AGACCAACTG CACACACTTT ACACTGATGC CCAGGCTAGC CCAGTCTAAA GGAAACACCA  660
ACATAGGAAG GGATGTGTGC AGGATTCACA AAAGATCTTT TCTACCCCCC GGAAAAACTA  720
AGTGGTGTGG TTTCGCTAAA CAGATTTTGC TAAGTACTTA AGCACTGCAG ATGCTTGAGT  780
AATATGCTCA TAAGTTCCTT TCTGATTTCA ATTACTGGGA AAATGTATAT ATGGATAGTA  840
GAAGGATGGC ATCCCATAAT AAAAGGCAGG CAGCCTAACC CTCACATGCA TTTTTCTCTC  900
CCTCTGTATA GGGTGACAGT GGAGGGCCTC TGGTTTGCTT CGAGAAGGAC AAATACATTT  960
TACAAGGAGT CACTTCTTGG GGTCTTGGCT GTGCACGCCC CAATAAGCCT GGTGTCTATG 1020
                                  XIX
TTCGTGTTTC AAGGTTTGTT ACTTGGATTG AGGGAGTGAT GAGAAATAAT TAATTGGACG 1080
GGAGACAGAG TGACGCACTG ACTCACCTAG AGGCTGGGAC GTGGGTAGGG ATTTAGCATG 1140
CTGGAAATAA CTGGCAGTAA TCAAACGAAG ACACTGTCCC CAGCTACCAG CTACGCCAAA 1200
CCTCGGCATT TTTTGTGTTA TTTTCTGACT GCTGGATTCT GTAGTAAGGT GACATAGCTA 1260
TGACATTTGT TAAAAATAAA CTCTGTACTT AACTTTGATT TGAGTAAATT TTGGTTTTGG 1320
TCTTCAACAT TTTCATGCTC TTTGTTCACC CCACCAATTT TAAATGGGCA GATGGGGGGA 1380
TTTAGCTGCT TTTGATAAGG AACAGCTGCA CAAAGGACTG AGCAGGCTGC AAGGTCACAG 1440
AGGGGAGAGC CAAGAAGTTG TCCACGCATT TACCTCATCA GCTAACGAGG GCTTGACATG 1500
CATTTTTACT GTCTTTATTC CTGACACTGA GATGAATGTT TTCAAAGCTG CAACATGCAT 1560
GGGGAGTCAT GCGAACCGAT TCTGTTATTG GGAATGAAAT CTGTCACCGA CTGCTTGACT 1620
```

FIGURE 2j

FIG. 2J

```
TGAGCCCAGG GGACACAGAG CAGAGAGCTG TATATGATGG AGTGAACCGG TCCATGGATG 1680
TGTAACACAA GACCAACTGA GAGTCTGAAT GTTATCCTGG GGCACACGTG AGTCTAGGAT 1740
TGGTGCCAAG AGCATGTAAA TGAACAACAA GCAAATATTG AAGGTGGACC ACTTATTTCC 1800
CATTGCTAAT TGCCTGCCCG GTTTTGAAAC AGTCTGCAGT ACACACGGTG ACAGGAGAAT 1860
GACCTGTGGG AGAGATACAT GTTTAGAAGG AAGAGAAAGG ACAAAGGCAC ACGTTTTACC 1920
ATTTAAAATA TTGTTACCAA ACAAAAATAT CCATTCAAAA TACAATTTAA CAATGCAACA 1980
GTCATCTTAC AGCAGAGAAA TGCAGAGAAA AGCAAAACTG CAAGTGACTG TGAATAAAGG 2040
GTGAATGTAG TCTCAAATCC TCAAAGAGCT GTGTTTATTT CATTGACAAA TAGATTATTT 2100
GTATCAA                                                          2107
```

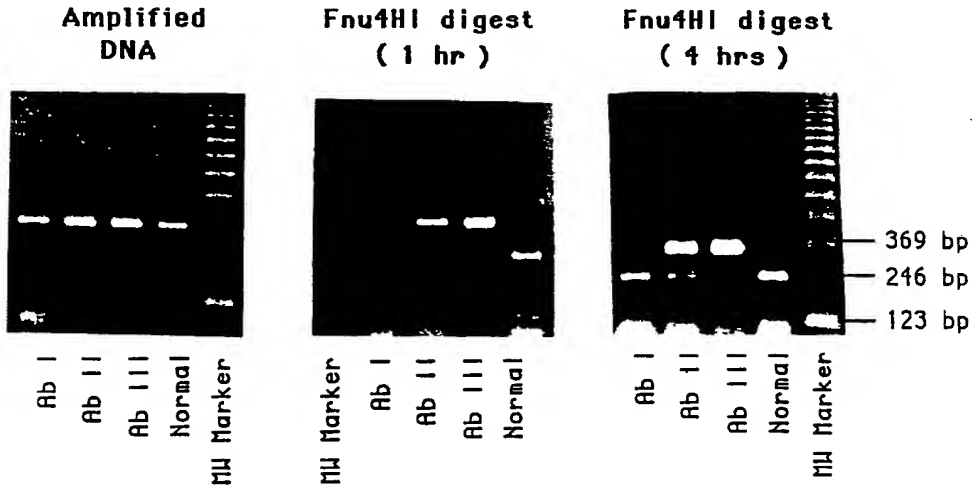
FIG. 2J CONT.

FIG.3

I.
INTRON H
EXON 14
INTRON H
V L I A H C
AAGTATCAATTTAACTAAAATTGAACTAAATCCICTT...50 bp...GGTGTTGACTGCTGCCCACTGC...180 bp...AAACTGCAGCAAAGACACAGTACAGATGATTTTGTGGCCCTGAATAAACTG
5'-AAGCTICAATTTAACTAAAATTGAACTAAAT-3'
HindIII (A39, 32mer)
HincII FnuHI
PstI
3'-TCATGTCTACTAAAACACCCGGACTTATTCGAG-5'
(1A, 33mer)
HindIII
II.
V L I A H C
TCTGCCTTTCTGTACAATGCAGCCAGAACAAAGTATCA....80 bp...GGTGTTGACTGTCGCCACTGC....165 bp...AGGCTTTCTAGAAAGAAACTGCAGAGGAAAGCACACAGTACAGATGATTTTGTGG
5'-TCTGGaaTTCTGTACAATGCAGCAGAACAAA-3'
EcoRI (10A, 31mer)
FnuHI
3'-TCTCCTTTCTGTGTCATGTCTACTTAAGCACC-5'
(11A, 32mer)
EcoRI

FIG. 4

Restriction Digest of the Amplified DNA



INTERNATIONAL SEARCH REPORT

International Application No PCT/US 89/02731

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) [6]

According to International Patent Classification (IPC) or to both National Classification and IPC

IPC : 4 C 12 Q 1/68, // C 07 H 21/04

II. FIELDS SEARCHED

Minimum Documentation Searched [7]

Classification System [8]	Classification Symbols
IPC 4	C 12 Q

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched [9]

III. DOCUMENTS CONSIDERED TO BE RELEVANT [9]

Category [10]	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No. [13]
A	Proc. Natl. Acad. Sci. USA, volume 85, January 1988, D.R. Engelke et al.: "Direct sequencing of enzymatically amplified human genomic DNA", pages 544-548 see the whole article --	1-4,18,19
A	EP, A, 0258017 (CETUS CORP.) 2 March 1988 see abstract; page 13, line 54 - page 21, line 39 --	1-4,18,19
A	EP, A, 0237362 (CETUS CORP.) 16 September 1987 see the whole document --	1-4,18,19
A	EP, A, 0256630 (HOWARD HUGHES MEDICAL INSITUTE) 24 February 1988 see the whole document --	1-4,18,19
A	WO, A, 84/01389 (MASSACHUSETTS INSTITUTE OF TECHNOLOGY) 12 April 1984 ./.	1-4,18,19

* Special categories of cited documents: [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
12th October 1989	20 NOV 1989
International Searching Authority	Signature of Authorized Officer
EUROPEAN PATENT OFFICE	T.K. WILLIS

III. DOCUMENTS CONSIDERED T BE RELEVANT (C NTINUED FR M THE SECOND SHEET)

Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No
	see abstract; claims 1-12 --	
A	FEBS Letters, volume 213, no. 2, March 1987, Elsevier Science Publishers B.V. (Biomedical Division), M. Forsgren et al.: "Molecular cloning and characterization of a full-length cDNA clone for human plasminogen", pages 254-260 cited in the application --	
A	Proc. Natl. Acad. Sci. USA, volume 79, October 1982, T. Miyata et al.: "Plasminogen Tochigi: inactive plasmin resulting from replacement of alanine-600 by threonine in the active site", pages 6132-6136 cited in the application	

ANNEX TO THE INTERNATIONAL SEARCH REPORT ON INTERNATIONAL PATENT APPLICATION NO. US 8902731 SA 29911

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report. The members are as contained in the European Patent Office EDP file on 15/11/89
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP-A- 0258017	02-03-88	AU-A- 7729887	19-05-88
		JP-A- 63102677	07-05-88
EP-A- 0237362	16-09-87	AU-A- 6996287	17-09-87
		JP-A- 62214355	21-09-87
EP-A- 0256630	24-02-88	AU-A- 7628187	04-02-88
		AU-A- 7653887	11-02-88
		JP-A- 63139194	10-06-88
WO-A- 8401389	12-04-84	US-A- 4535058	13-08-85
		EP-A- 0120958	10-10-84
		EP-A- 0241961	21-10-87
		EP-A- 0239162	30-09-87
		JP-T- 60500002	10-01-85
		US-A- 4786718	22-11-88